

AD _____

Award Number: DAMD17-98-1-8600

TITLE: A Micro-Simulation Model of the Benefits and Costs of
Prostate Cancer Screening and Treatment

PRINCIPAL INVESTIGATOR: David O. Meltzer, M.D., Ph.D.
Ibrahim Abdalla, M.D.

CONTRACTING ORGANIZATION: The University of Chicago
Chicago, Illinois 60637

REPORT DATE: January 2002

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20021024 005

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 074-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE January 2002	3. REPORT TYPE AND DATES COVERED Final (1 Jul 98 - 31 Dec 01)		
4. TITLE AND SUBTITLE A Micro-Simulation Model of the Benefits and Costs of Prostate Cancer Screening and Treatment		5. FUNDING NUMBERS DAMD17-98-1-8600		
6. AUTHOR(S) David O. Meltzer, M.D., Ph.D. Ibrahim Abdalla, M.D.				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The University of Chicago Chicago, Illinois 60637 E-Mail: dmeltzer@medicine.bsd.uchicago.edu		8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012		10. SPONSORING / MONITORING AGENCY REPORT NUMBER		
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 Words) This idea development award developed a micro-simulation model of prostate cancer screening and treatment. As described in our initial proposal, we developed an detailed model of the natural history, screening, and treatment of prostate cancer organized around a Markov model that incorporated true stage and grade of tumor, misclassification of tumors that generates discrepancies between clinical and true stage, and other innovative features that had been neglected in prior models. We also extended the model to account for persistent heterogeneity that would not be adequately modeled by usual Markov models. We also did a detailed structured literature review to estimate key parameters of our model and developed new techniques to estimate the parameters of the natural history model by fitting it to aggregate and individual-level data from the SEER program. Our baseline analysis suggests that screening and treatment will increase life expectancy but decrease quality-adjusted life expectancy. Potential gains are largest in men ages 50-70 and when undiagnosed cancer creates substantial anxiety. We published a paper based on some initial analyses of treatment patterns, have a main cost-effectiveness manuscript in progress, and received R01 funding from the National Cancer Institute to further extend this work.				
14. SUBJECT TERMS prostate cancer, cancer screening, micro-simulation model, markov model, tumor, SEER program			15. NUMBER OF PAGES 60	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

Table of Contents

Cover.....	Page 1
SF 298.....	2
Introduction.....	4
Body.....	4
Key Research Accomplishments.....	9
Reportable Outcomes.....	9
Conclusions.....	11
References.....	12
Appendices.....	15

Introduction:

There is great uncertainty about how to best screen for and treat prostate cancer. Randomized trials of screening and treatment are underway, but will take years to complete. In the meantime, patients and their doctors face difficult decisions. For these reasons, decision analysis has been widely used to study the costs and effectiveness of screening and treatment for prostate cancer. We developed a comprehensive model of the natural history, screening, and treatment of prostate cancer based on pathologic stage, grade of tumor, and other important aspects neglected by previous decision analysis. Model parameters are derived by extensive literature review and validated using aggregate and individual-level SEER data. We also extended the model to assess the effectiveness, costs, and cost-effectiveness of screening and treatment policies from age 50 to 99. Our baseline model suggests that screening and treatment improve non-discounted life expectancy for men ages 50-70 with more advanced tumors, but offer only modest increase for well-differentiated tumors. When quality of life weights were considered, quality-adjusted life expectancy was reduced. The more advanced sensitivity analysis is in progress with funding from the National Cancer Institute.

Body:

As specified in the **Statement of Work** section of our original proposal, all the goals are accomplished in line with the time schedule.

Overall, our model is comprised of three key features. The first is a model of the natural history of prostate cancer. This consists of a Markov process defined by a state vector that allows us to carry states of symptoms and stages from year to year and a transition matrix that describes the progression of prostate cancer. A decision tree is developed to model the probability that a man has prostate cancer detected in a year given the disease state that he carries from the beginning of the year (see Appendix 1). The second is the solid parameter justification. Model parameters are estimated by extensive literature review, second data analysis and mathematical modeling. We use maximum likelihood approach to validate the parameters using aggregate incidence and mortality data as well as individual level data. The third is estimation of cost- effectiveness with various screening policies and treatment policies.

We will report our progress in relation to the **Statement of Work** below.

Task 1. Review of model structure to reflect evolving innovation, Months 1-3.

- a) *Relevance given current knowledge of prostate biology (i.e. grade, p53, early metastases)*
- b) *Relevance given current screening and treatment technology (i.e. PSA density, % free PSA)*
- c) *Relevance for potential future technologies (i.e. PCR, Indium-111 labeled antibody to detect metastases, gene therapy)*

One important feature in our model is a state vector containing elements to carry from year to year. To determine what biological aspects are necessary to incorporate in the

state vector, we have conducted extensive literature review on the biology of prostate cancer. Although prostate cancer has often been characterized as a slow-growing tumor, the natural history of prostate cancer varies tremendously based on tumor grade and stage.^{1,2} Therefore, we have included prostate cancer grade in our state vector that we carry from year to year. Our comprehensive literature review has helped us determine the probabilities that a man transitions among states.

Although the identification of p53, a tumor suppressor gene, has been found to be one important advance in prostate,³ its utilization in standard practice is premature to the point. Before its prognostic use is confirmed by future studies, we have not included information concerning p53 into our decision analysis. However, our model is flexibly programmed to permit the inclusion of new development concerning p53.

Prostate specific antigen(PSA) testing has diagnostic and prognostic value in practice. PSA test remains as the most relevant technology for detecting prostate cancer. It is also increasingly popular in clinical staging of prostate cancer tumors. Therefore, we have incorporated PSA level detectability as a element in the state vector of our model.

However, the benefits offered by PSA density remain to be further explored.⁴ Likewise, the data is still too limited to determine whether free PSA is clinically useful.^{5,6,7} As such, we have not formally incorporated such measurement techniques into the model, but our model is flexible enough to include those techniques if their diagnostic or prognostic value are confirmed.

Our comprehensive literature review indicates none of the new therapies listed above (PCR, Indium-111 labeled antibody to detect metastases, gene therapy) have come into widespread use as diagnostic or treatment tools for men with prostate cancer. However, The flexibility of our model allows us to include the effects of new diagnostic and therapies on prostate cancer detection, treatment, and outcomes.

Task 2. Programming of model to reflect revised structure, Months 1-6

- a) Revision of decision tree if needed*
- b) Revision of natural history model if needed*
- c) Programming of costs*
- d) Programming of benefits and quality of life adjustments*

We have refined our decision tree substantially to make our model more accurate compared to the original one in our proposal. For example, in the original model, we assume a given percentage of men would have DRE detectable prostate cancer. However, because men who do not have DRE detectable prostate cancer are less likely to have it the next year, the original approach would have overestimated the cumulative number of men who have DRE detectable prostate cancer. Therefore, the likelihood that men would have rectal detectable cancer should be conditional on whether or not symptoms were detected in the previous year. To address the complication of progression of prostate cancer, we have created a large matrix with 528 states to carry over year. The states incorporate combinations of cancer symptom(two states), screening status(two states), BPH symptoms(two states), metastatic detectability(two states), rectal and PSA detectability (six states), having had a TURP(two states), true state (five states), and true grade(three

states). The probability of moving into each of the states is conditional on the state in previous year.

We have designed our model to be flexible so that we can allow for the changes in costs, benefits, and quality of life. We have conducted a number of sensitivity analyses concerning quality of life, relecting the difference of utility obtained through specific treatment. Orginially, we incorporate the anxiety into our model as a main complication of watchful waiting treatment. The results indicate the potential gains can be large if the patients experience substantial anxiety.

Task 3. Review of literature to estimate model parameters, Months 4-18

- a) Identification of parameters of interest*
- b) Collection of articles relevant for parameters of interest*
- c) Review and analysis of articles*
- d) Preparation of documentation of literature review*

We have refined the decision tree model to identify the parameters of importance. We estimate each model parameter via one of the following approaches: 1) comprehensive literature review; 2) analysis of primary or secondary data; 3) mathematical modeling. For example, to estimate the probability that a biopsy will be positive in the presence of stage A cancer, we design a geometric model of a needle transversing a hypothetical prostate of a standard volume encountering a tumor of specified volume present in the prostate. We also validate the parameter with post-mortem studies of biopsy yield in men examined for prostate cancer on autopsy. The complete discussion on our parameter justification is in the paper "Patterns of prostate cancer treatment by clinical stage and age in the United States," which has been published in the American Journal of Public Health.

Our extensive literature review covers many carefully performed analysis of prostate cancer screening and treatment.^{8,9,10,11,12,13,14,15,16,17,18,19,20} Although most of the papers have many strengths, all of them have major flaws undermining their validity and relevance. The first and most important flaw is that those studies fail to model prostate cancer incidence as a dynamic process that may be affected by previous screening patterns. The second major limitation of existing model is that they fail to distinguish either between clinical stage and pathological stage, or between cancers of differing stage and grade. The existing literature has also minor flaws, including ignoring the sensitivity/specificity of screening tests, failing to consider quality of life, and a variety of data-related issues.

Conclusively, based on the existing models, we have developed more complicated model to account for extra key aspects of the natural history, screening, and treatment of prostate cancer. As highlighted in our annual report, those factors are to model:

- (1). Tumor heterogeneity and progression rates by grade.
- (2). Misclassification of tumor stage due to the discrepancy between clinical and pathologic stage.
- (3). The ability of screening and treatment to affect the prevalence of prostate cancer.
- (4). The effect of BPH and other prostate symptoms and their management on the detection of prostate cancer.
- (5). The potential effects of screening and treatment of quality of life, as measured by quality-adjusted life expectancy.

Task 4: Sensitivity Analysis: Months 6-18

- a) *All parameters in model*
- b) *One and multi-way sensitivity analysis*
- c) *Assess effects on:*
 - 1) *Costs*
 - 2) *Effectiveness*
 - 3) *Cost-effectiveness*

We have completed the sensitivity analyses regarding to the parameter estimates derived from literature review, meta-analyses, and secondary data analysis. Our model allows for three levels of parameters, which are the best estimate, the upper and lower boundaries of the parameters. But some parameters, such as quality of life weights, have no analytical bounds. Therefore, we use the range of quality of life weights in the literature for sensitivity analysis. Finally, we employ the maximum likelihood approach to optimize our decision model. However, it is a very time-consuming task which can take months to achieve the optimization. If one parameter or assumption is modified, the whole process has to run over again. The final optimized result is presented in Appendix 2. More detailed discussion is in Task 5.

We have conducted extensive research on cost-effectiveness analysis. We adopt the expected utility maximization approach to perform the sensitivity analysis. Generally speaking, utility assessment reflects the strength of the preference or the degree of abhorrence for the potential outcome of interest. Our work focuses on determining patient utilities for different states with prostate cancer and treatment complications for the cost-effectiveness analysis. The output of our research is to date a series of publications on cost-effectiveness analysis, as listed in the **Reportable Outcomes** section below., and more publications are in progress.

Another original contribution to the existing literature is to incorporate future costs into the cost-effectiveness analysis. We have published a number of papers^{21,22,23}, indicating the importance of including future medical and non-medical costs in cost-effectiveness analysis. Simply put, we add costs associated with non-prostate cancer related medical expenditures and non-medical consumption net of earnings to total costs accumulated in each years of life. Theoretically, the inclusion of future costs will most adversely affect screening and treatment that have relative small effects on quality-adjusted life expectancy compared to life expectancy. Therefore, the cost-effectiveness analysis with future costs will favor the interventions that minimize complications of treatment for prostate cancer.

Task 5: Estimation of Transition Rates Using SEER and Watchful Waiting Data, Months 12-24

- a) *Programming of estimation techniques*
- b) *Data cleaning*
- c) *Estimation*
- d) *Sensitivity Analyses*

As mentioned in Task 3, the most important contribution of our decision model is

to simulate the prostate cancer incidence as a dynamic process that is neglected by most existing models. The key parameters of our model are the transition probabilities between stage, grade and death. Although we can estimate most parameters of interest via literature review or data analysis, there is no information available concerning the transition probability. Therefore, we adopt the curve fitting technique to estimate the transition matrix. Specifically, the constrained optimization is introduced to reduce the sum of squares difference between our predicted aggregate incidence and mortality data and SEER data. As can be seen in the Appendix 2, the predicted mortality results and clinical stage results fit the SEER results very well, though not perfectly. Because of the time limitation and computing power limitation, it is extremely difficult to completely reduce the sum of squares to zero. We have received R01 funding from the National Cancer Institute to continue our optimization in the aggregate level. In the meantime, we will extend the optimization to the individual level, i.e., optimizing the maximum likelihood function constructed by the individual data. We will determine the analytic probability that a person in SEER will have an observed detection pattern, treatment pattern, and mortality pattern according to our decision model.

Task 6: Preparation of Papers Reporting Cost-Effectiveness Analyses, Months 18-30

- a) Evaluation of established existing technologies*
 - 1) Treatment, esp. of special patient subgroups*
 - 2) Rectal*
 - 3) PSA*
 - 4) Differing screening ages and intervals*
- b) Evaluations of innovative technologies and management strategies*
 - 1) Percent Free-PSA*
 - 2) Detection of micrometastases*
 - 3) Other technologies*

We have completed preliminary cost-effectiveness analysis based on treatment and age primarily. The detailed results are presented in Appendix 3.

Table 1a-3a in Appendix 3 provide the baseline cost-effectiveness results of age-specific treatments for clinically localized well, moderate and poor grade cancers respectively. The baseline analysis uses a discount rate of 3% and incorporates the baseline QOL weights of impotence, incontinence, urethral stricture, bowel dysfunction and metastasis to compute the quality-adjusted life-years (QALYs). There are some noticeable increases in non-discounted life years with treatments (radiation or surgery) over watchful waiting especially at younger ages over all grades of cancer. Moreover, the cost-effectiveness ratio of cost per discounted life years for treatments over watchful waiting is cost-effective by the \$100K/Life Years criteria at younger ages over all grades of cancer. However, once QOL weights are included in the calculations, all treatments become dominated by watchful waiting at all ages and all grades of cancer. The results stress the importance of patients' perception of side effects as an important factor in choosing treatment.

The results are found to be highly sensitive to psychological anxiousness of the patient detected with prostate cancer. No literature could be identified that studies the effect of such anxiety on the quality of life of patients. Moreover, this anxiety is expected to change with treatment if the cancer could be eliminated. We assigned a new QOL

weight to patient with anxiety. If patients are detected with prostate cancer their baseline QOL is equated to this anxiety QOL in the absence of any other side effects. If they are treated with surgery we assume that the cancer could be removed with surety and the patient's QOL will become 1 (normal) in the absence of any other side effects. In case of radiation, though the cancer may be fully cured there is no way to ascertain this for sure and hence the patient's QOL is given a weight = $(1 + \text{Anxiety QOL})/2$ in the absence of any other side effects.

Table 1b -3b shows the sensitivity analysis of anxiety QOL on the cost-effectiveness results. Modest increase in anxiety (or modest decrease in anxiety QOL) leads to favorable cost-effectiveness ratios for radiation and surgery. Such effects appear to be true over all ages and all grades of cancer, though younger patients with more advanced grade of cancer seem to benefit most from treatments if they suffer from anxiety.

The R01 grant from the National Cancer Institute will allow us to extend the cost-effectiveness under various screening policies, innovative technologies, and management strategies.

Key Research Accomplishments

- Created more accurate model concerning the natural history of prostate cancer and a more accurate decision tree concerning the treatment of prostate cancer.
- Identified relevant natural history, treatment, and outcome parameters for inclusion in our natural history and decision tree models.
- Determined the model parameter probabilities using literature reviews, meta-analyses, and primary data analysis.
- Completed preliminary cost-effectiveness analyses that indicate that treatment is cost-effective in terms of life years.
- Completed preliminary cost-effectiveness analyses that indicate that the cost-effective ratio is sensitive to QALYs resulting from treatment decisions at younger ages, and the anxiety QOL at all ages.
- Published a peer-reviewed journal paper describing prostate cancer treatment rates across clinical stages.
- Published a peer-reviewed journal paper describing the theoretical basis for addressing uncertainty in cost-effectiveness analysis
- Have almost completed a major paper examining the cost-effectiveness of prostate cancer treatment. Already presented in abstract form at the Society for Medical Decision Making.
- Received R01 funding from the National Cancer Institute to continue this work to final analysis of cost-effectiveness

(8) Reportable Outcomes

Grants and Fellowships:

*Cost-Effectiveness of Prostate Cancer Screening and Treatment. National Cancer Institute, 09/01/01- 08/31/03, \$1,142,433. R01-CA92443.

*John M. Olin Foundation Faculty Fellowship, 7/1/99/10/1/00, \$108,917.

Published Articles:

*Meltzer D, Egleston BL, Abdalla I. Patterns of Prostate Cancer Treatment by Clinical Stage and Age. *American Journal of Public Health*. 91(1): 126-128, 2001.

*Meltzer D. Addressing Uncertainty in Medical Cost-Effectiveness Analysis: Implications of Expected Utility Maximization for Methods to Perform Sensitivity Analysis and the Use of Cost-Effectiveness Analysis to Set Priorities for Medical Research. *Journal of Health Economics*, 20(1):109-129, 2001.

*Meltzer D, Johannesson M. Inconsistencies in the 'Societal Perspective' on Costs of the Panel on Cost-Effectiveness in Health and Medicine. *Medical Decision Making*, 19:371-377, 1999

*Meltzer D, Johannesson M. On the Role of Theory in Cost-Effectiveness Analysis: A Response To Garber, Russell, and Weinstein. *Medical Decision Making*, 19:383-384, 1999

Published Abstracts

*Meltzer D, Hazen G, Johannesson M, Abdalla I, Chang R, Elstein A, Schwartz A. Effect of Future Costs on the Cost-Effectiveness of Life Extension and Quality of Life Improvement Among the Elderly. *Medical Decision Making* 18: 475, October 1998.

*Meltzer D, Polonsky T. Do Quality-Adjusted Life Years Reflect Patient Preferences? Validation Using Revealed Preference for Intensive Treatment of Insulin-Dependent Diabetes Mellitus. *Medical Decision Making* 18: 459, October 1998.

*Meltzer D, Basu A, Egleston B. Structure and Early Results of a Prostate Cancer Decision Model. *Medical Decision Making*. In press.

Presentations:

"Implications of Expected Utility Maximization for Assessing Cost-Effectiveness under Uncertainty: Methods for Sensitivity Analysis and Implications for the Use of Cost-Effectiveness Analysis to Set Priorities for Medical Research". National Bureau of Economic Research Summer Institute, Boston, MA, July 1998.

"Effect of Future Costs on the Cost-Effectiveness of Life Extension and Quality of Life Improvement Among the Elderly", Society for Medical Decision Making Annual Meeting, Plenary Address, Boston, MA, October 1998.

- "Accounting for Future Costs from a Societal Perspective in Cost-Effectiveness Analysis": Plenary Address, International Health Economics Association Annual Meeting, Rotterdam, Netherlands, June 1999.
- "Can Burden of Illness Measures Aid Priority-setting at NIH?". Invited Presentation to Dr. Harold Varmus and panel of experts convened at NIH in Response to IOM Report on Scientific Priorities at NIH, June 1999.
- "Measuring Quality-Adjusted Years of Life due to Different Diseases". National Bureau of Economic Research Summer Institute, Boston, MA, July 1999.
- "Do People Consider Financial Effects in Answering Quality of Life Questions?" Plenary Abstract, Society for Medical Decision Making Annual Meeting, Reno, NV, October 1999.
- "Do QALYs Measure What We Want?" Harvard School of Public Health, November 1999.
- "Can Medical Cost-Effectiveness Analysis Identify the Value of Research?" Conference on the Economic Value of Medical Research organized by the Lasker Foundation, Washington, DC, December 1999.
- "Theoretical Foundations of Medical Cost-Effectiveness Analysis", Northeastern Illinois Chapter of the American Statistical Association, Chicago, IL, March 2000.
- "Accounting for Future Costs in Cost-Effectiveness of Prostate Cancer Screening and Treatment". Robert Wood Johnson Generalist Physician Faculty Scholars Program, Annual Meeting, December 2000.
- "The Value of Information in Cost-Effectiveness Analysis, Application to Prostate Cancer Screening and Treatment" International Health Economics Association, (iHEA), Annual Meeting, York, England, July 2001.
- "Structure and Early Results of a Prostate Cancer Decision Model." Society of Medical Decision Making 23rd Annual Meeting. San Diego, CA October 2001.

Conclusions

We have conducted extensive literature review on the natural history, screening, treatment, and cost-effectiveness of prostate cancer. To remedy the flaws in most existing models, we have developed the most complicated decision analytic model of prostate cancer to date. In particular our model is an improvement on other models in that it accounts for:

- (1) Tumor heterogeneity and progression rates by grade,

- (2) Misclassification of tumor stage due to the discrepancy between clinical and pathologic stage.
- (3) The ability of screening and treatment to affect the prevalence of prostate cancer.
- (4) The effect of benign prostatic hypertrophy and other prostate symptoms and their management on the detection of prostate cancer.
- (5) The potential effects of screening and treatment on quality of life, as measured by quality adjusted life expectancy.

We have estimated all parameters in the model via literature review, meta-analysis, secondary data analysis, and mathematical modeling. We use the data from SEER program to validate the transition probabilities across stage, grade and morality.

We have completed the baseline cost-effectiveness of treatment of prostate cancer. Our results indicate treatments offer increases in non-discounted life years especially at younger age below 70 over all grades. However, when quality of life weights were added, quality-adjusted life expectancy was reduced compared with watchful waiting. The results are sensitive to the patient's anxiety over the treatment. Serious anxiousness of the patient can favor the radiation and surgery over watchful waiting. We will continue this important work and conduct the extensive sensitivity analysis under R01 grant from the National Cancer Institute.

References

- ¹ Chodak GW, Thisted RA, Gerber GS, Johansson JE, Adofsson J, Jones GW, Chisholm GD, Moskovitz B, Livne PM, Warner J. Results of conservative management of clinically localized prostate cancer. *NEJM* 1994; 330:242-248.
- ² Johansson JE, Amadio HO, Johansson S, Bergstrom R, Adami HO. Fifteen-year survival in prostate cancer, a prospective, population-based study in Sweden. *JAMA* 1997; 277:467-471
- ³ deVere White RW, Gumerlock PH, Chi S-G, Meyers FJ. p53 tumor suppressor gene abnormalities are frequent in human prostate tissues. *Journal of Urology* 1993; 149:376A.
- ⁴ Mettlin C, Littrup PJ, Kane RA, Murphy GP, Lee F, Chesley A, Badalament R, Mostofi FK. Relative sensitivity and specificity of serum prostate specific antigen (PSA) level compared with age-referenced PSA, PSA density, and PSA change: Data from the American Cancer Society National Prostate Cancer Detection Project. *Cancer* 1994; 74:1615-1620.
- ⁵ Prestigiacomo AF, Lilja H, Pettersson K, Wolfert RL, Stamey TA. A comparison of the free fraction of serum prostate specific antigen in men with benign and cancerous prostates: The best case scenario. *Journal of Urology* 1996; 156:350-354.
- ⁶ Pannek J, Rittenhouse HG, Chan DW, Epstein JI, Walsh PC, Partin AW. The use of percent free prostate specific antigen for staging clinically localized prostate cancer. *Journal of Urology* 1998; 159:1238-1242.
- ⁷ Catalona WJ, Partin AW, Slawin KM, Brawer MK, Flanigan RC, Patel A, Richie JP, deKernion JB, Walsh PC, Scardino PT, Lange PH, Subong EN, Parson RE, Gasior GH, Loveland KG, Southwick PC. Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic disease: A prospective multicenter trial. *JAMA* 1998; 279:1542-1547.

- ⁸ Love RR, Fryback DG. A Cost-Effectiveness Analysis of Screening for Carcinoma of the Prostate by Digital Examination. *Medical Decision Making* 1985; 5(3):263-278.
- ⁹ Carlsson P, Pedersen KV, Varenhorst E. Costs and benefits of early detection of prostatic cancer. *Health Policy* 1990; 16:241-253.
- ¹⁰ Mold JW, Holtgrave DR, Bisonni RS, Marley DS, Wright RA, Spann SJ. The Evaluation and Treatment of Men with Asymptomatic Prostate Nodules in Primary Care: A Decision Analysis. *The Journal of Family Practice* 1992; 34(5):561-568.
- ¹¹ Fleming C, Wasson JH, Albertsen PC, Barry MJ, Wennberg JE. A Decision Analysis of Alternative Treatment Strategies for Clinically Localized Prostate Cancer. *JAMA* 1993; 269(20):2650-2658.
- ¹² Littrup PJ, Goodman AC, Mettlin CJ. The Benefit and Cost of Prostate Cancer Early Detection. *CA-Cancer Journal for Clinicians* 1993; 43:134-149.
- ¹³ Littrup PJ, Kane RA, Mettlin CJ, Murphy GP, Lee F, Toi A, Badalament R, Babaian R. Cost-Effective Prostate Cancer Detection: Reduction of Low-Yield Biopsies. *Cancer* 1994; 74(12):3146-3158.
- ¹⁴ Krahn MD, Mahoney JE, Eckman MH, Trachtenberg J, Pauker SG, Detsky AS. Screening for prostate cancer: A decision analytic view. *JAMA* 1994; 272(10): 773-80.
- ¹⁵ Cowen ME, Chartrand M, Weitzel WF. A Markov Model of the Natural History of Prostate Cancer. *Journal of Clinical Epidemiology* 1994; 47(1):3-21.
- ¹⁶ Optenberg SA, Thompson IM. Economics of Screening for Carcinoma of the Prostate. *Urologic Clinics of North America* 1990; 17(4): 719-737.
- ¹⁷ Benoit RM, Naslund MJ. An Economic Rationale for Prostate Cancer Screening. *Urology* 1994; 44(6):795-803.
- ¹⁸ Cantor SB, Spann SJ, Volk RJ, Cardenas MP, Warren MM. Prostate Cancer Screening: A Decision Analysis. *The Journal of Family Practice* 1995; 41(1):33-41.
- ¹⁹ Barry MJ, Fleming C, Coley CM, Wasson JH, Fahs MC, Oesterling JE. Should Medicare provide reimbursement for prostate-specific antigen testing for early detection of prostate cancer? - Part IV: Estimating the risks and benefits of an early detection program. *Urology* 1995; 46(4):445-461.

²⁰ Kattan MW, Cowen ME, Miles BJ. A decision analysis for treatment of clinically localized prostate cancer. *Journal of General Internal Medicine* 1997; 12:299-305.

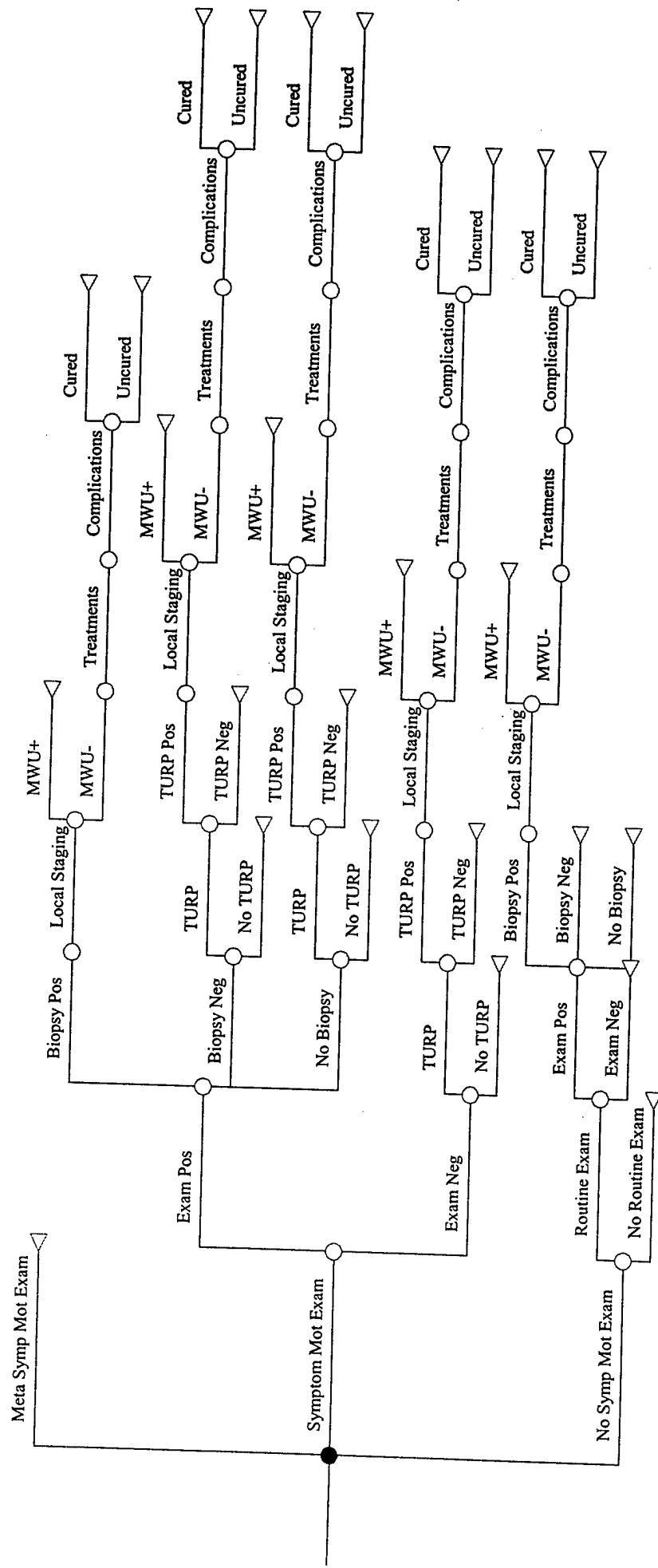
²¹ Johannesson M, Meltzer D, O'Connor R. Incorporating Future Costs in Medical Cost-Effectiveness Analysis: Implications for the Cost-Effectiveness of the Treatment of Hypertension. *Medical Decision Making* 1997; 17:382-389.

²² Meltzer D. Accounting for future costs in medical cost-effectiveness analysis. *Journal of Health Economics* 1997; 16:33-64.

²³ Meltzer DO, Egleston B, Stoffel D, Dasbach E. The effect of future costs on the cost-effectiveness of medical interventions among young adults: The example of intensive therapy for type-1 diabetes. *Medical Care* 2000; 38:679-685.

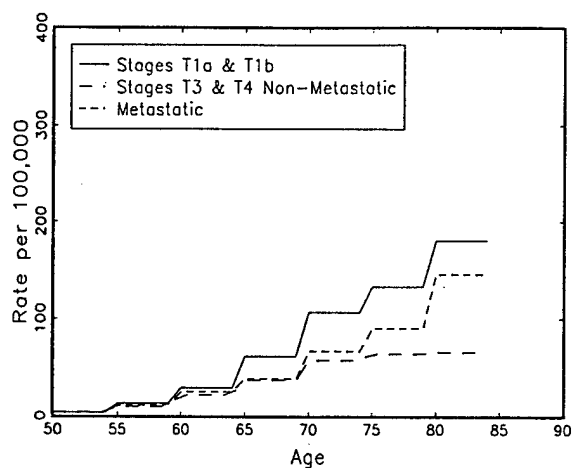
Appendices

- Appendix 1. Decision Tree Model.
- Appendix 2. The Model Optimization Results Compared with SEER Data
- Appendix 3. Cost-effectiveness Analysis of Treatment of Prostate Cancer
- Appendix 4. Meltzer D, Egleston BL, Abdalla I. Patterns of Prostate Cancer Treatment by Clinical Stage and Age. *American Journal of Public Health*. 91(1): 126-128, 2001.
- Appendix 5. Meltzer D. Addressing Uncertainty in Medical Cost-Effectiveness Analysis: Implications of Expected Utility Maximization for Methods to Perform Sensitivity Analysis and the Use of Cost-Effectiveness Analysis to Set Priorities for Medical Research. *Journal of Health Economics*, 20(1):109-129, 2001.
- Appendix 6. Meltzer D, Johannesson M. Inconsistencies in the 'Societal Perspective' on Costs of the Panel on Cost-Effectiveness in Health and Medicine. *Medical Decision Making*, 19:371-377, 1999
- Appendix 7. Meltzer D, Johannesson M. On the Role of Theory in Cost-Effectiveness Analysis: A Response To Garber, Russell, and Weinstein. *Medical Decision Making*, 19:383-384, 1999
- Appendix 8. Meltzer D, Hazen G, Johannesson M, Abdalla I, Chang R, Elstein A, Schwartz A. Effect of Future Costs on the Cost-Effectiveness of Life Extension and Quality of Life Improvement Among the Elderly. *Medical Decision Making* 18: 475, October 1998.
- Appendix 9. Meltzer D, Polonsky T. Do Quality-Adjusted Life Years Reflect Patient Preferences? Validation Using Revealed Preference for Intensive Treatment of Insulin-Dependent Diabetes Mellitus. *Medical Decision Making* 18: 459, October 1998.
- Appendix 10. Meltzer D, Basu A, Egleston B. Structure and Early Results of a Prostate Cancer Decision Model. *Medical Decision Making*. In press.

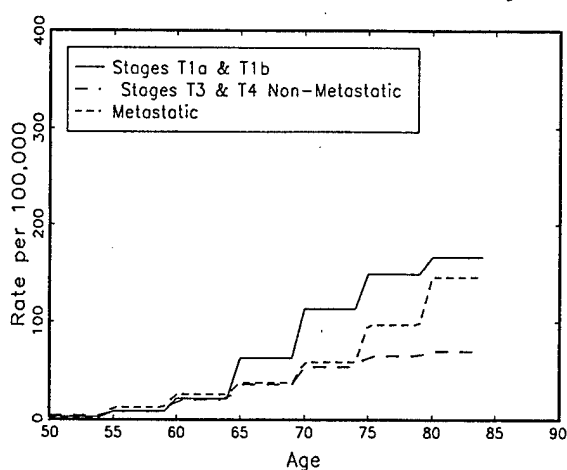


Appendix 2: The Model Optimization Results Compared with SEER Data

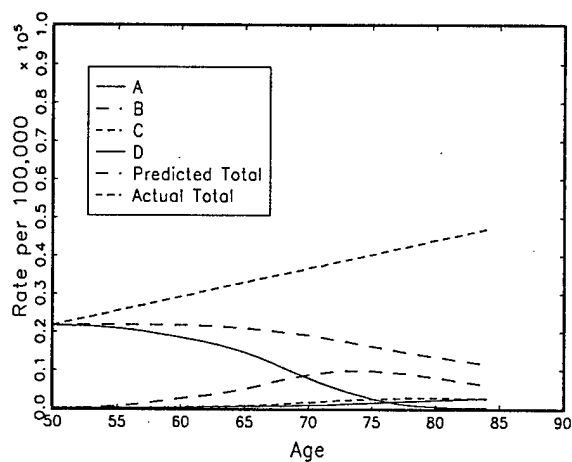
1995 Actual SEER Incidence - Clinical Stage



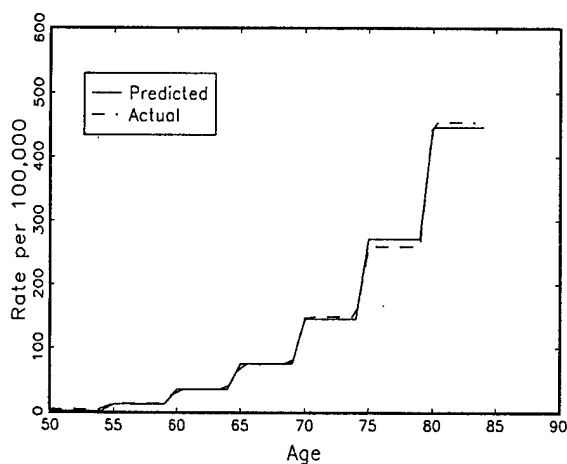
1990s Predicted Incidence - Clinical Stage



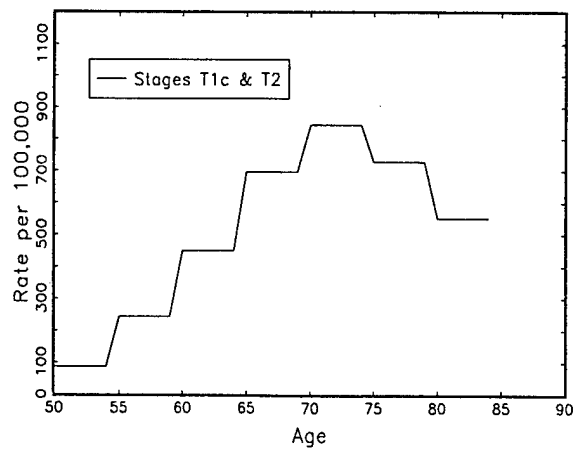
Predicted Prevalence - True StageABCD



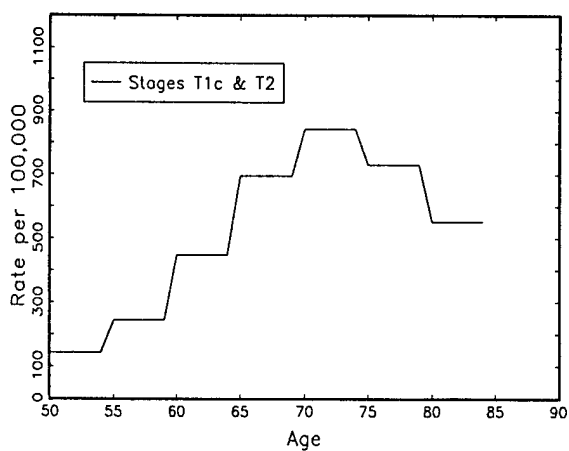
1990s Cancer Death Rates



1995 Actual SEER Incidence - Clinical Stage



1990s Predicted Incidence - Clinical Stage



Appendix 3: Cost-effectiveness Analysis of Treatment of Prostate Cancer

Table 1a: Cost Effectiveness Analysis of age-specific treatments for well differentiated and clinically localized cancer.

Age	Treatments	Discounted Cost*	Incremental Cost	Non-Disc Life Years	Discounted Life Years	Discounted QALYs	Incremental QALYs	ICER* (1000\$/QALY)	ICER* (vs WW)
55 yrs	Watchful Waiting	5483	-	21.82	13.39	13.32	-	-	-
	Surgery	34515	-	22.53	13.63	9.84	-3.47	Dominated	-8.4
	Radiation	35125	-	22.64	13.69	11.45	-1.87	Dominated	-15.9
65 yrs	Watchful Waiting	5993	-	14.83	7.41	7.37	-	-	-
	Surgery	39044	-	15.17	7.52	5.43	-1.94	Dominated	-17.0
	Radiation	39733	-	15.25	7.56	6.32	-1.05	Dominated	-32.0
75 yrs	Watchful Waiting	4612	-	9.24	3.73	3.71	-	-	-
	Surgery	31935	-	9.26	3.73	2.69	-1.02	Dominated	-26.7
	Radiation	32502	-	9.32	3.75	3.13	-0.58	Dominated	-48.1

* Without future costs. + in 1000/QALY

** QALY include baseline values of QOL for Impotence, Incontinence, Urethral Stricture, Bowel Dysfunction & Metastasis

Table 1b: Sensitivity Analysis with Anxiety QOL: CEA of age-specific treatments for well-differentiated and clinically localized cancer.

Anxiety QOL ==>		0.6		0.7		0.8		0.9		1.0	
AGE	Treatments	Discounted Cost*	Incremental Cost	QALY dQ	ICER	QALY dQ	ICER	QALY dQ	ICER	QALY dQ	ICER
55 yrs	Watchful Waiting			7.99	-	9.32	-	10.65	-	11.99	-
	Surgery			+1.85	15.7	+0.52	55.7	-0.81	Dom	-2.14	Dom
	Radiation			+0.06	10.3	+0.45	1.4	+0.02	1359.7	-0.92	Dom
65 yrs	Watchful Waiting			4.42	-	5.16	-	5.90	-	6.63	-
	Surgery			+1.01	32.8	+0.27	122.3	-0.47	Dom	-1.20	Dom
	Radiation			+0.03	19.7	+0.25	2.8	-0.001	Dom	-0.53	Dom
75 yrs	Watchful Waiting			2.23	-	2.60	-	2.97	-	3.34	-
	Surgery			+0.46	59.0	+0.09	296.7	-0.28	Dom	-0.65	Dom
	Radiation			+0.02	30.8	+0.12	4.6	-0.05	Dom	-0.31	Dom

* Without future costs. ** QALY include baseline values of QOL for Impotence, Incontinence, Urethral Stricture, Bowel Dysfunction & Metastasis.

Table 2: Cost Effectiveness Analysis for age-specific treatment for moderately-differentiated and clinically localized cancer.

Age	Treatments	Discounted Cost*	Incremental Cost	Non-Disc Life Years	Discounted Life Years	Discounted QALYs**	Incremental QALYs	ICER*	ICER* (vs WW)
55 yrs	Watchful Waiting	6463		19.88	12.49	12.28		-	-
	Surgery	27630		21.61	13.18	9.48	-2.80	Dominated	-7.6
	Radiation	28106		21.71	13.24	11.02	-1.26	Dominated	-17.2
65 yrs	Watchful Waiting	4784		13.53	6.91	6.76		-	-
	Surgery	20181		14.46	7.23	5.19	-1.56	Dominated	-9.9
	Radiation	20527		14.53	7.27	6.04	-0.72	Dominated	-21.9
75 yrs	Watchful Waiting	3320		8.71	2.39	3.48		-	-
	Surgery	17389		8.85	2.39	2.57	-0.91	Dominated	-15.4
	Radiation	17689		8.90	2.40	2.99	-0.50	Dominated	-29.0

* Without future costs. + in 1000/QALY

** QALY include baseline values of QOL for Impotence, Incontinence, Urethral Stricture, Bowel Dysfunction & Metastasis

Table 2b: Sensitivity Analysis with Anxiety QOL: CEA of age-specific treatments for moderately differentiated and clinically localized cancer.

Anxiety QOL ==>		0.6				0.7				0.8				0.9				1.0			
AGE	Treatments	Discounted Cost*	Incremental Cost	QALY dQ	ICER	QALY dQ	ICER	QALY dQ	ICER	QALY dQ	ICER	QALY dQ	ICER	QALY dQ	ICER	QALY dQ	ICER	QALY dQ	ICER	QALY dQ	ICER
55 yrs	Watchful Waiting	6463		7.37	-	8.59	-	9.82	-	11.05	-	12.28	-	11.05	-	12.28	-	11.05	-	12.28	-
	Surgery	27630		+2.02	10.5	+0.84	25.1	-0.37	Dom	-1.58	Dom	-2.80	Dom	-1.58	Dom	-2.80	Dom	-1.58	Dom	-2.80	Dom
	Radiation	28106		+0.15	3.2	+0.47	1.0	+0.45	47.7	-0.40	Dom	-1.26	Dom	-0.40	Dom	-1.26	Dom	-0.40	Dom	-1.26	Dom
65 yrs	Watchful Waiting	4784		4.05	-	4.73	-	5.41	-	6.08	-	6.76	-	6.08	-	6.76	-	6.08	-	6.76	-
	Surgery	20181		+0.99	15.5	+0.40	38.7	-0.25	Dom	-0.91	Dom	-1.56	Dom	-0.91	Dom	-1.56	Dom	-0.91	Dom	-1.56	Dom
	Radiation	20527		+0.18	2.0	+0.30	1.2	+0.22	70.0	-0.25	Dom	-0.72	Dom	-0.25	Dom	-0.72	Dom	-0.25	Dom	-0.72	Dom
75 yrs	Watchful Waiting	3320		2.09	-	2.44	-	2.79	-	3.13	-	3.48	-	3.13	-	3.48	-	3.13	-	3.48	-
	Surgery	17389		+0.33	42.7	+0.06	223.3	-0.26	Dom	-0.58	Dom	-0.91	Dom	-0.58	Dom	-0.91	Dom	-0.58	Dom	-0.91	Dom
	Radiation	17689		+0.16	1.8	+0.18	1.6	-0.001	Dom	-0.25	Dom	-0.50	Dom	-0.25	Dom	-0.50	Dom	-0.25	Dom	-0.50	Dom

* Without future costs. ** QALY include baseline values of QOL for Impotence, Incontinence, Urethral Stricture, Bowel Dysfunction & Metastasis.

Table 3: Cost Effectiveness Analysis for age-specific treatment for poorly differentiated and clinically localized cancer.

Age	Treatments	Discounted Cost*	Incremental Cost	Non-Disc Life Years	Discounted Life Years	Discounted QALYs**	Incremental QALYs	ICER*	ICER* (vs WW)
55 yrs	Watchful Waiting	8047		16.48	10.75	10.38	-	-	-
	Surgery	23435		19.21	11.93	8.52	-1.86	Dominated	-8.3
	Radiation	23815		19.30	11.98	9.89	-0.49	Dominated	-32.2
65 yrs	Watchful Waiting	5646		10.85	5.73	5.43	-	-	-
	Surgery	12613		11.78	6.08	4.28	-1.15	Dominated	-6.1
	Radiation	12795		11.84	6.11	4.95	-0.48	Dominated	-14.1
75 yrs	Watchful Waiting	2466		7.15	2.99	2.80	-	-	-
	Surgery	6225		7.24	3.01	2.09	-0.71	Dominated	-7.0
	Radiation	6315		7.28	3.03	2.41	-0.39	Dominated	-13.1

* Without future costs. .+ in 1000/QALY

** QALY include baseline values of QOL for Impotence, Incontinence, Urethral Stricture, Bowel Dysfunction & Metastasis

Table 3b: Sensitivity Analysis with Anxiety QOL: CEA of age-specific treatments for poorly differentiated and clinically localized cancer.

AGE	Treatments	Anxiety QOL ==>									
		Discounted Cost*	Incremental Cost	0.6		0.7		0.8		0.9	
				QALY dQ	ICER	QALY dQ	ICER	QALY dQ	ICER	QALY dQ	ICER
55 yrs	Watchful Waiting	8047		6.23	-	7.26	-	8.30	-	9.34	-
	Surgery	23435		+2.07	7.4	+1.15	13.3	+0.16	97.6	-0.85	Dom
	Radiation	23815		+0.26	1.5	+0.47	0.8	+0.76	0.5	+0.21	73.5
65 yrs	Watchful Waiting	5646		3.26	-	3.80	-	4.34	-	4.88	-
	Surgery	12613		+0.72	9.7	+0.34	20.4	-0.15	Dom	-0.65	Dom
	Radiation	12795		+0.31	0.6	+0.31	0.6	+0.27	26.1	-0.10	Dom
75 yrs	Watchful Waiting	2466		1.68	-	1.96	-	2.24	-	2.52	-
	Surgery	6225		+0.16	31.1	+0.02	332.3	-0.22	Dom	-0.47	Dom
	Radiation	6315		+0.24	0.5	+0.19	0.6	+0.01	694.4	-0.19	Dom

* Without future costs. ** QALY include baseline values of QOL for Impotence, Incontinence, Urethral Stricture, Bowel Dysfunction & Metastasis.

Appendix 4:

Patterns of Prostate Cancer Treatment by Clinical Stage and Age

Meltzer D, Eggleston B, Abdalla I.

American Journal of Public Health

91(1): 126-128, 2001

ABSTRACT

Objectives. This study analyzed prostate cancer treatment rates by age and clinical stage and contrasted these with rates by most accurate stage.

Methods. We determined surgery and radiation rates by most accurate and clinical stage by using 1996 Surveillance, Epidemiology, and End Results data.

Results. Treatment rates by clinical stage vs best stage differ significantly. For example, surgery rates for stages B, C, and D are 37%, 78%, and 13% by most accurate stage but 33%, 6%, and 1% by clinical stage. Treatment patterns by clinical stage vary substantially by age.

Conclusions. Treatment patterns should be described by clinical stage rather than most accurate stage, and they vary by age. (*Am J Public Health*. 2001; 91:126-128)

Patterns of Prostate Cancer Treatment by Clinical Stage and Age

David Meltzer, MD, PhD, Brian Eggleston, MPP, and Ibrahim Abdalla, MD

Data on variations in prostate cancer treatment patterns may provide insight into the effects of new screening technologies such as prostate-specific antigen or into the acceptance of new therapeutic approaches such as brachytherapy, or they may raise questions about access to or appropriateness of treatment.¹⁻⁹ Since treatment is generally tailored to the patient's age and tumor stage, data on treatment patterns are most useful when treatment rates are reported by age and stage.

Unfortunately, even when treatment rates are reported by stage, they are often based on the most accurate stage rather than the clinical stage at the time treatment decisions were made.^{6,9,10} The most accurate stage is determined by starting with the clinical stage and replacing it with the pathologic stage if surgery is performed. Thus, assessments of tumor stage based on clinical staging tests, such as digital rectal examination, are superseded by pathologic information obtained from surgery. Since pathologic staging of prostate cancer following prostatectomy often results in a different stage than was determined by clinical staging,¹¹ reports of cancer treatment based on most accurate stage may not represent how prostate cancer is actually treated upon its clinical presentation.

This study used 1997 Surveillance, Epidemiology, and End Results (SEER)¹² data to calculate treatment rates by age according to clinical stage, and it contrasted these rates with estimates of treatment rates according to most accurate stage. The SEER program, which collects data on cancer incidence, treatment, and mortality from cancer registries that now cover approximately 14% of the US population, is believed to be reasonably representative of the United States.¹³

Methods

We defined clinical stage as follows: A, clinically localized and nonpalpable on rectal examination; B, clinically localized but palpable; C, palpable with clinical evidence of local extension beyond the prostate; D, lymph node involvement or distant metastases. SEER also reports the most accurate stage of cancer, incorporating any information available from prostatectomy. In these cases, localized disease without extension is classified automatically as stage B, because staging guidelines used by SEER categorize all organ-confined tumors as

stage B.¹⁴ In addition, we classified patients with lymph nodes that tested positive for cancer as being in pathologic stage D.

We categorized treatment as surgery (prostatectomy), radiation (including brachytherapy), combined surgery and radiation, or neither treatment. We calculated treatment rates by age on the basis of both most accurate and clinical stage.

The sample comprises 22 578 men older than 50 years with prostate cancer. Of these, 131 were excluded because they had cystoprostatectomies and 4865 were excluded owing to missing or inconsistent data. Those excluded did not differ substantially from those included by age, race, or pathologic stage where information was available for comparison.

Results

Treatment Rates

Overall, 30% of patients receive surgery, 32% receive radiation, 1% receive combined therapy, and 37% receive neither surgery nor radiation. When patients are combined across all ages and staging is defined by most accurate stage, no stage A patients receive prostatectomies (because SEER does not define a pathologic stage A), while 37% of stage B and 78% of stage C patients receive prostatectomies. When the data are analyzed by clinical stage, surgery rates are slightly higher than 30% for stages A and B and are only 6% for stage C and 1% for stage D. Radiation rates by most accurate stage for stages A, B, C, and D are 41%, 36%, 19%, and 16%, while rates by clinical stage are 27%, 38%, 54%, and 16%.

Treatment rates vary substantially by age, whether by most accurate stage (Figure 1) or clinical stage (Figure 2), with the fraction of

David Meltzer is with the Section of General Internal Medicine, Department of Economics, and Harris Graduate School of Public Policy Studies, Brian Eggleston is with the Section of General Internal Medicine, and Ibrahim Abdalla is with the Department of Radiation and Cellular Oncology and the Department of Urology, University of Chicago, Chicago, Ill.

Requests for reprints should be sent to David Meltzer, MD, PhD, 5841 S Maryland Ave, MC 2007, Chicago, IL 60637 (e-mail: dmeltzer@medicine.bsd.uchicago.edu).

This brief was accepted July 20, 2000.

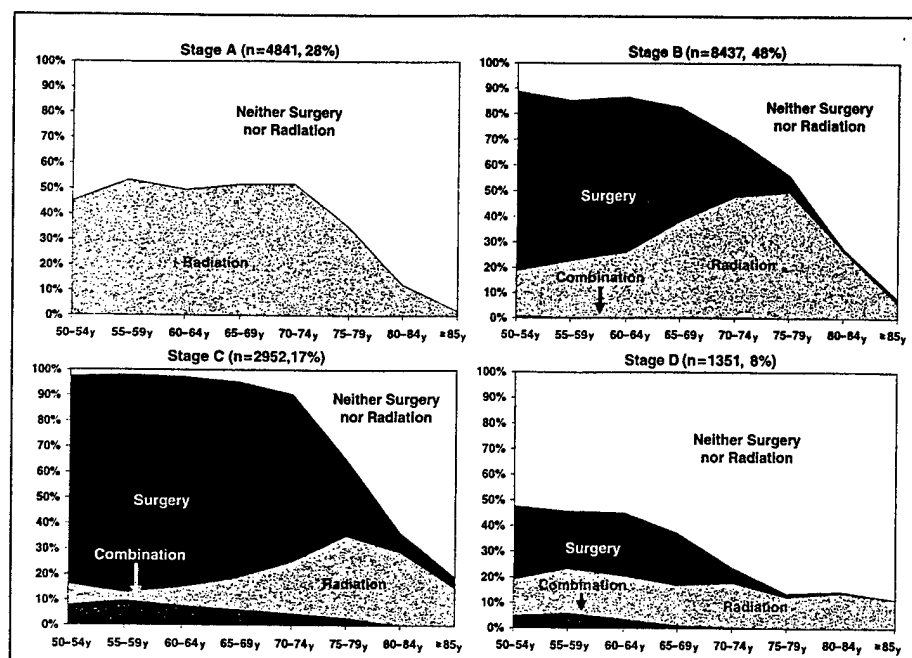


FIGURE 1—Prostatectomy and radiation treatment rates for prostate cancer according to age and stage defined by most accurate information.

accurate stage also substantially overstates the proportion of men choosing radiation for clinical stage A cancer and understates the proportion choosing radiation for clinical stage C cancer.

Treatment Patterns by Age

Most men younger than 65 years with stage A or B cancer choose surgery, but surgery rates fall rapidly after 70 years of age. Radiation rates rise with age until peaking in the 70s, when they begin to decline. The overall pattern by age is similar for stage C patients, although surgical rates are generally lower and radiation rates are higher. Most men with clinical stage D cancer receive neither surgery nor radiation, although surgical rates are higher if stage D is defined by most accurate stage.

Pathologic Restaging

Because SEER does not identify a pathologic stage A, the cancer stage of all men with clinical stage A cancer is reclassified after surgery; this includes one man who was found not to have cancer after prostatectomy. Cancer stage was reclassified for nearly half of the men with clinical stage B cancer and approximately one third of the men with clinical stage C cancer. Seventy-seven percent of stage C cancers by most accurate stage were the result of restaging following prostatectomy.

Discussion

Treatment Patterns by Clinical vs Most Accurate Stage

Our results confirm that substantial upstaging occurs after surgery, causing treatment rates based on most accurate stage to substantially underestimate surgery rates for patients with clinical A and B disease while overestimating surgery rates for stage C disease. This probably explains why previous analyses of treatment rates based on most accurate stage found substantially lower rates of surgery for stage A cancer and higher rates of surgery for stage C cancer than we find here.⁶ These findings demonstrate how important it is that cancer databases retain information on both clinical and pathologic states, so that treatment rates can be reported on the basis of clinical stage.

Treatment Patterns by Age

In contrast to prior studies that did not report rates by combinations of age and clinical stage, our results show that surgery is by far the predominant method of treating stage A

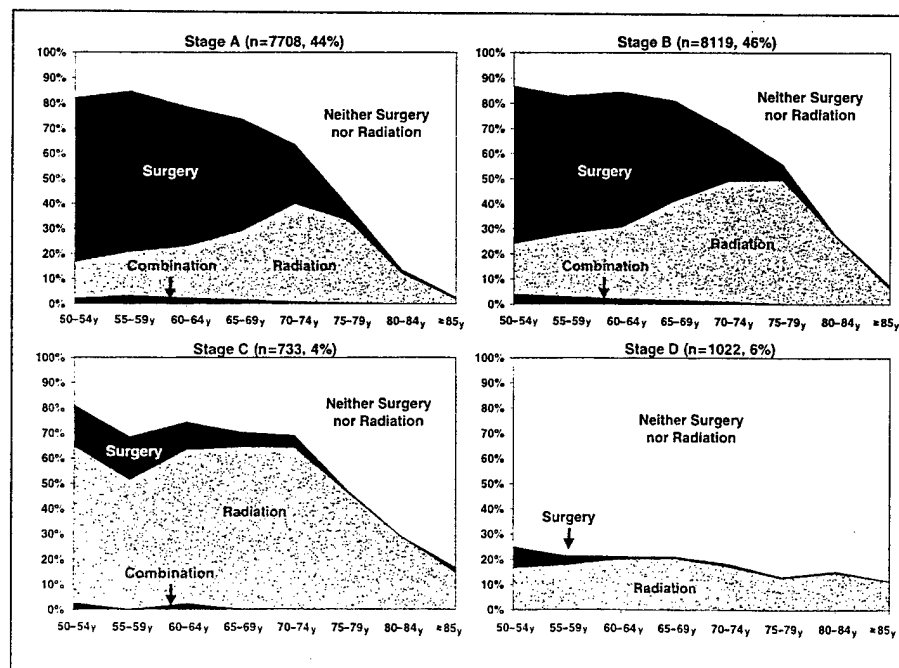


FIGURE 2—Prostatectomy and radiation treatment rates for prostate cancer according to age and stage defined by clinical diagnosis.

patients receiving neither surgery nor radiation almost uniformly increasing by age, but especially after 75 years. By most accurate stage, 71% of the youngest men with stage B and 89% of the youngest men with stage C cancer have prostatectomies. By clinical stage, prostatectomy rates are about 67% for the youngest men with stage A and B cancer but only 19%

and 8% for the youngest men with stage C and D cancer. The overstatement of surgery rates by most accurate vs clinical stage is even greater for men aged 55 to 74 years. Because using most accurate stage understates the fraction of men with clinical stage A cancer choosing surgery and overstates the fraction with clinical stage C cancer choosing surgery, using most

and B prostate cancer among younger men, with radiation predominant among somewhat older men.

A striking finding is that nearly a fifth of younger men with clinical stage C cancer are treated surgically, despite the decreased likelihood that they will benefit from surgery. With staging based on the "most accurate" information, this relationship might have been explained by patients with clinically observed A or B disease who were then found to have stage C disease. Our findings suggest this is not the case, and they intensify the need to understand whether such surgery is justified.

Limitations

SEER holds advantages, including the comprehensiveness of its population-based sample, continuity over time, and attention to quality control, but it also has limitations. First, SEER relies on medical and pathologic records that can be incomplete.¹⁵ Additionally, SEER does not report whether lymph node involvement was found before surgery or as part of pelvic lymphadenectomy performed immediately before a prostatectomy that was consequently aborted. This lack of data may cause the stage and treatment of some patients to be misclassified, a possibility that suggests the value of collecting information on the mechanism and timing of assessments of lymph node involvement. □

Contributors

All authors contributed to the conceptualization and execution of the study and to the writing of the paper. In addition, D. Meltzer identified the need for analysis of treatment patterns by clinical stage and directed the study design and execution, B. Eggleston performed the data analysis, and I. Abdalla provided clinical expertise.

Acknowledgments

This work was supported by the Department of Defense Prostate Cancer Research Program (DAMD 17-98-1-8600), the National Institute on Aging (K12 AG00488-08), and the Robert Wood Johnson Generalist Physician Faculty Scholars Program.

References

1. Harlan L, Brawley O, Pommerenke F, Wali P, Kramer B. Geographic, age, and racial variation in the treatment of local/regional carcinoma of the prostate. *J Clin Oncol*. 1995;13:93-100.
2. Lu-Yao G, Greenberg ER. Changes in prostate cancer incidence and treatment in USA. *Lancet*. 1994;343:251-254.
3. Mettlin CJ, Murphy G. The National Cancer Data Base report on prostate cancer. *Cancer*. 1994;74:1640-1648.
4. Polednak AP. Geographic variation in the treatment of prostate cancer in Connecticut. *Int J Technol Assess Health Care*. 1993;9:304-310.
5. Severson RK, Montie JE, Porter AT, Demers RY. Recent trends in incidence and treatment of prostate cancer among elderly men. *J Natl Cancer Inst*. 1995;87:532-534.

6. Mettlin CJ, Murphy GP, Cunningham MP, Menck HR. National Cancer Data Base report on race, age, and region variations in prostate cancer treatment. *Cancer*. 1997;80:1261-1266.
7. Lu-Yao G, McLerran D, Wasson J, Wennberg JE. An assessment of radical prostatectomy: time trends, geographic variation, and outcomes. *JAMA*. 1993;269:2633-2636.
8. Klabunde CN, Potosky AL, Harlan LC, Kramer BS. Trends and black/white differences in treatment for nonmetastatic prostate cancer. *Med Care*. 1998;36:1337-1348.
9. Mettlin CJ, Murphy GP, Rosenthal DS, Menck HR. The National Cancer Data Base report on prostate carcinoma after the peak in incidence rates in the US. *Cancer*. 1998;83:1679-1684.
10. Mettlin CJ, Murphy GP, Ho R, Menck HR. The National Cancer Data Base report on longitudinal observations on prostate cancer. *Cancer*. 1996;77:2162-2166.
11. Partin AW, Kattan MW, Subong EN, et al. Combination of prostate-specific antigen, clinical stage, and Gleason Score to predict pathological stage of localized prostate cancer. *JAMA*. 1997; 277:1445-1451.
12. National Cancer Institute. *SEER Cancer Incidence Public-Use Database, 1973-1995, August 1999 Submission* [CD-ROM]. Bethesda, Md: National Institutes of Health; April, 2000.
13. National Cancer Institute. SEER program. Available at: <http://www-seer.ims.nci.nih.gov>. Accessed October 12, 2000.
14. Fleming ID, Cooper JS, Henson DE, et al. *AJCC Cancer Staging Manual*. 5th ed. Philadelphia, Pa: Lippincott-Raven; 1997:219-224.
15. Fritz A, Ries L, eds. *SEER Program Code Manual*. Bethesda, Md: National Cancer Institute; 1998.

Appendix 5:

**Addressing Uncertainty in Medical Cost-Effectiveness Analysis: Implications of
Expected Utility Maximization for Methods to Perform Sensitivity Analysis and the
Use of Cost-Effectiveness Analysis to Set Priorities for Research**

David Meltzer

Journal of Health Economics

20(1): 109-129, 2001



Journal of Health Economics 20 (2001) 109–129

JOURNAL OF
**HEALTH
ECONOMICS**

www.elsevier.nl/locate/econbase

Addressing uncertainty in medical
cost-effectiveness analysis
Implications of expected utility maximization for
methods to perform sensitivity analysis and
the use of cost-effectiveness analysis to
set priorities for medical research

David Meltzer*

*Section of General Internal Medicine, Harris Graduate School of Public Policy Studies, Department of
Economics, University of Chicago, 5841 S. Maryland Avenue MC 2007, Chicago, IL 60637, USA*

Received 1 April 1999; accepted 29 August 2000

Abstract

This paper examines the objectives for performing sensitivity analysis in medical cost-effectiveness analysis and the implications of expected utility maximization for methods to perform such analyses. The analysis suggests specific approaches for optimal decision making under uncertainty and specifying such decisions for subgroups based on the ratio of expected costs to expected benefits, and for valuing research using value of information calculations. Though ideal value of information calculations may be difficult, certain approaches with less stringent data requirements may bound the value of information. These approaches suggest methods by which the vast cost-effectiveness literature may help inform priorities for medical research. © 2001 Elsevier Science B.V. All rights reserved.

JEL classification: I18; D61; O32

Keywords: Cost-effectiveness analysis; Uncertainty; Sensitivity analysis; Health care research; Value of information; Welfare economics

* Tel.: +1-773-702-0836; fax: +1-773-834-2238.
E-mail address: dmeltzer@medicine.bsd.uchicago.edu (D. Meltzer).

1. Introduction

Despite some recent slowing in the growth of health care costs in the US, health care costs have risen substantially over the past several decades and are likely to continue rising (Smith et al., 1998). This appears to be largely due to the growth of new technology (Fuchs, 1990; Newhouse, 1992). While improvements in health are highly valued (Cutler and Richardson, 1997; Murphy and Topel, 1998), evidence from diverse methodological perspectives suggests that many technologies may have little value at the margin (Eddy, 1990; Brook et al., 1983; McClellan et al., 1994). Cost-effectiveness analysis and other methods for medical technology assessment have arisen to attempt to address this important problem.

One of the main challenges faced by medical cost-effectiveness analysis has been the question of how to perform these analyses in the presence of uncertainty about the benefits and costs of medical interventions. The uncertainty of primary interest in this regard is uncertainty in population level outcomes, although uncertainty in outcomes at the individual level may be present simultaneously. This uncertainty in population level outcomes may result either from limited evidence from clinical trials or the need to extrapolate based on the results of clinical trials using decision analysis and its associated uncertainties in the structure and parameters of decision models. This uncertainty concerning the benefits and costs of medical interventions has motivated much interest in sensitivity analysis within medical cost-effectiveness analysis.

Yet though there have been many proposals about how to address uncertainty in cost-effectiveness analysis, there has been relatively little discussion of the objectives for performing sensitivity analysis. Without a clear understanding of these objectives, it is difficult to know by what criterion to assess the merits of the many alternative approaches to sensitivity analysis. Thus, the lack of clarity concerning the objectives for sensitivity analysis is an important reason for the continuing ambiguity about how to address uncertainty in cost-effectiveness analysis.

This paper attempts to identify the objectives for sensitivity analysis within cost-effectiveness analysis and to develop methods suited to reaching those objectives. The primary objectives of sensitivity analysis are argued to be: (1) to help a decision maker make the best decision in the presence of uncertainty, (2) to identify the sources of uncertainty to guide decisions for individuals or subgroups with characteristics that differ from a base case, and (3) to set priorities for the collection of additional information. This paper studies these problems by examining the implications of an expected utility maximization model for the optimal choice of medical interventions when there is uncertainty about the costs and benefits of those interventions. The results indicate that if the objective is to maximize expected utility given available information — as is implicit, for example, in the maximization of quality-adjusted life expectancy — and if financial risk is effectively diversified through either public or private insurance, then the optimal decision is determined by the ratio of the expected cost divided by the expected benefit. Other assumptions about preferences or insurance will yield other conclusions about how to account for uncertainty (Mullahy, 1997), but also would require different models for cost-effectiveness in the absence of uncertainty at the population level. These findings also have implications for sensitivity analyses done for other purposes. If the objective of sensitivity analysis is to guide decisions for subgroups that differ from the base case, then the ratio of expected

costs to expected benefits for that subgroup is the appropriate criterion. If the objective of sensitivity analysis is to set priorities for the acquisition of additional information, then the incremental increase in expected utility with additional information is the appropriate measure of benefit. Though such ideal value of information calculations may be difficult to perform, other approaches to sensitivity analysis with less stringent data requirements may provide bounds on the value of information. Together, these approaches suggest a theoretically grounded approach by which the tools of medical cost-effectiveness analysis can be used to help set priorities for medical research. Following these approaches, it may be possible to draw upon the vast literature on the cost-effectiveness of specific medical interventions (Elixhauser et al., 1998) to address crucial needs for more systematic ways to set priorities for medical research. After active discussion between Congress, the Administration, and the leadership of the National Institutes of Health (NIH) over the value of and priorities for Federal funding of biomedical research, the need for such systematic approaches to identify priorities for research at the NIH was recently highlighted in a report of the Institute of Medicine (IOM, 1998).

Section 2 discusses the objectives of sensitivity analysis. Section 3 discusses the primary methods currently used to perform sensitivity analysis. Section 4 uses an expected utility maximization model to derive methods for optimal decision making in the context of uncertainty about population outcomes. Section 5 extends the basic results of Section 4 to encompass uncertainty at the individual level. Section 6 uses the model to derive methods for sensitivity analysis to guide decisions for individuals or subgroups that differ from a base case. Section 7 applies these principles to a stylized decision concerning a medical treatment of uncertain benefit. Section 8 uses the model to derive methods to use sensitivity analyses to inform priorities for the collection of additional information to guide decision making, including approaches to bound value of information calculations with limited information. Section 9 applies these ideas to a stylized model of the decision whether to treat prostate cancer and discusses some challenges in implementing these approaches to set priorities for research. Section 10 concludes.

2. Objectives for sensitivity analysis

In order to begin to assess methods to account for uncertainty in cost-effectiveness analysis, it is essential to consider the objectives in performing sensitivity analyses. Although not all of these objectives may be relevant in every application, the objectives appear to fall into three broad categories: (1) to help a decision maker make the best decision in the presence of uncertainty about costs and effectiveness, (2) to identify the sources of uncertainty to guide decisions for individuals or groups with characteristics that differ from a base case, and (3) to set priorities for the collection of additional information.

2.1. *Decision making under uncertainty about cost and effectiveness*

This is probably the most common reason that sensitivity analysis is performed in medical cost-effectiveness analysis, and arises because the scientific literature often does not provide precise information concerning effectiveness or costs. For example, the efficacy of an immunization or the frequency and cost of complications may not be known with

confidence. Nevertheless, patients must decide whether they want the immunization and public and private insurers must decide whether they will cover it. Thus, having a mechanism to help guide decision making when the costs and benefits of a medical intervention are uncertain is important.

2.2. Decision making for individuals or subgroups that differ from a base case

Though not frequently stated as a motivation for sensitivity analysis, developing insight into decisions faced by individuals or subgroups is also a common motivation for performing sensitivity analysis in medical cost–effectiveness analysis. For example, a cost–effectiveness analysis for immunization of a population would likely consider the average risk of acquiring an infection in the absence of immunization. However, an analyst examining the cost–effectiveness of immunization for an individual or group with a known risk factor for acquiring some infection would want to reflect that higher-than-average risk.

2.3. Priority-setting for the collection of additional information

When the conclusions of a cost–effectiveness analysis are altered by parameter values that cannot be ruled out based on the literature, the collection of additional information concerning those parameters may be justified. Though in practice it is not frequently done, sensitivity analysis can be used to identify parameters that may change the results of a decision analysis and those parameters may then be studied more intensively. A few studies have used this approach to determine the value of sample size for clinical trials (Claxton and Posnett, 1996; Hornberger, 1998), or to perform sensitivity analysis in a decision model by calculating the expected value of perfect information concerning specific parameters of the model (Felli and Hazen, 1998).

Although these three motivations for performing sensitivity analysis are clearly distinct, papers in the literature commonly do not distinguish among them in their discussion of the sensitivity analysis. This is important because different methods for sensitivity analysis may be better suited to different objectives. This is discussed further below.

3. Methods for sensitivity analysis

Before attempting to derive methods for performing sensitivity analysis, it is useful to discuss the existing methods. The oldest and most commonly used forms of sensitivity analysis are univariate sensitivity analyses. Following these approaches, analysts begin with the mean or modal values of all the probabilities in their analysis and use those to calculate the costs and benefits for a “base case” analysis. The parameters are then varied individually across a range of possible outcomes to see how the cost–effectiveness of an intervention changes. In some instances, the parameter values are varied over the range of all possible values, while in other cases they are varied across confidence intervals that are drawn from the medical literature.

A major advantage of one-way sensitivity analyses is that they permit the analyst to identify the effects of individual parameters on the analysis. Another advantage of one-way sensitivity analyses is that the results can be easily calculated and reported. However, there

are also a number of significant shortcomings of these approaches. First, they do not clearly delineate either what range of parameter values to consider or what to do when some of those possible parameter values would change the optimal decision. For example, consider again the case of a vaccination. Its probability of providing immunity is logically constrained to a number between 0 and 1. High and low estimates in the literature might be 0.98 and 0.60. The best study might predict a protection rate of 0.92 with a 95% confidence interval of 0.89–0.96. Which of these are we to choose in setting the range of parameters? If we choose the broadest range, it may be impossible to pin down the costs and benefits with sufficient precision to determine whether the intervention is worthwhile. If we use a 95% confidence interval and find a benefit throughout the range, the potential for an immense harm that could occur if the true value of the parameter falls outside that range would fail to be recognized. Even if we find that the optimal decision changes for a parameter value at the upper end of the 95% confidence interval, it is not clear how that should change the decision we should make. If the welfare benefits over the majority of the interval are large, and any welfare loss at an extreme of the confidence interval is modest, it is not clear that the negative result at the extreme should have much influence on the decision made. Threshold analyses — which identify the parameter values at which an analysis crosses a cost-effectiveness threshold — are subject to the same criticism for failing to reflect the magnitude of the effect of the parameter on costs and outcomes and therefore the significance of the fact that the cost-effectiveness ratio crosses some threshold for some parameter values.

Another concern with one-way sensitivity analyses is that they may be misleading if the results obtained by varying a parameter depend on the level of other parameters in the model. This has motivated multi-way sensitivity analyses in which parameters are varied simultaneously across plausible or likely ranges. These analyses are subject to all the concerns described above concerning one-way sensitivity analyses, as well as some additional problems. One problem with these approaches is that the number of sensitivity analyses that must be performed rises exponentially with the number of parameters. Another problem is that assumptions about one parameter often have implications for assumptions about other parameters. For example, assumptions about the natural history of untreated disease may have implications for the history of disease under treatment. This has motivated efforts to examine the joint distribution of the parameters in a model. This approach, along with a similar population-based sampling approach to estimating costs, effectiveness and cost-effectiveness ratios, sometimes termed stochastic cost-effectiveness analysis, appears to be receiving increasing attention in the field (O'Brien et al., 1994; Gold et al., 1996; Polsky et al., 1997). However, these analyses still do not address the question of the optimal decision in the presence of uncertainty because they do not suggest what to do when the set of possible costs and outcomes include ones that would make the cost-effectiveness ratio fail to meet the chosen threshold for cost-effectiveness.

Furthermore, there are a set of issues related to the calculation of cost-effectiveness ratios as ratios, and the relationship between those ratios and resource allocation. The ratio issue is important because benefits and costs will usually not be distributed independently since changes in parameter values will often influence both simultaneously and because ratios and their expectations are undefined when the benefits are, or may be, zero. Likewise, cost-effectiveness ratios may have very different meanings depending on the signs of costs and benefits so that merely calculating averages or even confidence intervals for

cost-effectiveness ratios would not generally be meaningful (Stinnett and Paltiel, 1997). One creative approach to these issues is to reformulate cost-effectiveness analyses in terms of Net Health Benefits (Stinnett and Mullahy, 1998), in which both costs and benefits are expressed in the common denominator of years of life saved. While free of some of the complications associated with estimating cost-effectiveness ratios, the utility of the Net Health Benefit approach is diminished by the fact it does not allow easy comparisons with results from traditional cost-effectiveness analyses that rely on cost-effectiveness ratios, and is dependent on assumptions about the valuation of improvements in health. A related approach with similar concerns is to convert health benefits into a monetary value, as is done in cost-benefit analysis (Tambour et al., 1998).

In assessing these methods, it is interesting to note that while all of them appear to have some significance for the objectives described above, none of them are explicitly linked to those objectives. As discussed above, this lack of clarity concerning the objectives for sensitivity analysis is an important reason for the continuing ambiguity concerning methods to account for uncertainty in medical cost-effectiveness analysis. The next two sections use an expected utility maximization model to attempt to develop an approach to assess the importance of uncertainty about parameter values in order to make an optimal decision under uncertainty. The sections that follow then examine the adaptation of that approach to address the other two common objectives of sensitivity analysis — the determination of cost-effectiveness for individuals or subgroups and the identification of areas where the collection of additional information would be of value.

4. A deterministic model of health outcomes with uncertainty about effectiveness

In this simple case, we assume that there is uncertainty about the effectiveness ($\theta \in \Theta$, with pdf $p(\theta)$) of providing m units of medical care (for example, blood pressure checks per year), but that the outcome of that medical care given θ is certain. By making this assumption, we abstract from the problem of uncertainty in outcome for an individual, and focus instead on uncertainty for a “representative consumer” assumed to be identical to all other individuals, so that there is no heterogeneity in the population. We return to these issues of individual level uncertainty and heterogeneity in Sections 5 and 6, however.

To capture the possibility that effectiveness may affect both the costs and benefits of an intervention, we allow both utility (U) and the costs of the medical care (c) to depend directly on θ so $c = c(m, \theta)$. This allows the cost of m units of medical care to be uncertain, as it might be, for example, if it is not known how much those blood pressure checks and resulting treatments would cost. In addition, utility is assumed to depend on non-medical consumption (x) and medical expenditure, so $U = U(m, \theta, x(\theta))$. Here x is written as $x(\theta)$ to denote the fact that x will vary with θ for any m to satisfy the budget constraint $c(m, \theta) + x(\theta) - I = 0$ for each level of effectiveness. To model cost-effectiveness, we assume that people maximize expected utility¹ and take the example of a representative

¹ While individual preferences may in fact be inconsistent with expected utility maximization, QALYs implicitly assume that people maximize expected utility. While relaxing this assumption might be desirable, doing so would therefore involve a substantial reformulation of the way in which health benefits are assessed even in the absence of uncertainty. This is discussed further in Section 10.

consumer who maximizes expected utility subject to budget constraint conditional on each level of effectiveness:

$$\max_m \int p(\theta) U(m, \theta, x(\theta)) d\theta \text{ such that } c(m, \theta) + x(\theta) - I = 0 \text{ for all } \theta. \quad (1)$$

Rewriting this as a Lagrange multiplier problem with $\lambda(\theta)$ as the multiplier for the budget constraint at each level of θ , and multiplying each $\lambda(\theta)$ by $p(\theta)$ without loss of generality yields

$$\max_{m, \lambda(\theta)} \int p(\theta) U(m, \theta, x(\theta)) d\theta + \int \lambda(\theta) p(\theta) [I - c(m, \theta) - x(\theta)] d\theta. \quad (2)$$

This generates a first-order condition for medical expenditure which is

$$\int p(\theta) \frac{\partial U(m, \theta, x(\theta))}{\partial m} d\theta + \int \lambda(\theta) p(\theta) \frac{\partial c(m, \theta)}{\partial m} d\theta = 0. \quad (3)$$

This implies that investment in a medical intervention should occur to the point at which its expected marginal benefit (utility) equals the expected value of the marginal-utility-of-income-weighted marginal cost. Allowing the marginal utility of income to depend on θ reflects the possibility that, either because of changes in the utility function or costs with θ , income might have a greater or lesser marginal utility.

For an individual, these effects of uncertainty about the costs and effectiveness of medical interventions on the marginal utility of income are clearly plausible and potentially important. If someone has hip replacement for arthritis at age 55 and then suffers a severe complication, is forced into early retirement, and requires around-the-clock care, both their utility and medical costs will be directly affected and their marginal utility of income could change substantially. In a population, however, such effects are far less compelling because insurance can equate the marginal utility of income across health states unless an intervention leads to an extraordinarily large change in either population health or costs. Thinking from a population perspective in which most extremely expensive medical interventions affect a relatively small number of persons and most common medical interventions are relatively modest in cost, it is much less likely that the (aggregate) marginal utility of income will change substantially with uncertainty about the costs or benefits of a single intervention.² If this is the case, then $\lim \lambda(\theta) \rightarrow \lambda$ and the first-order condition for medical expenditures converges to

$$\int p(\theta) \frac{\partial U(m, \theta, x(\theta))}{\partial m} d\theta + \int \lambda p(\theta) \frac{\partial c(m, \theta)}{\partial m} d\theta = 0, \quad (4)$$

which implies that the cost-effectiveness ratio is

$$\frac{\int p(\theta) (\partial c(m, \theta) / \partial m) d\theta}{\int p(\theta) (\partial U(m, x(\theta)) / \partial m) d\theta} = \frac{1}{\lambda}. \quad (5)$$

² Note that even if changes in health status led to substantial changes in income or the need for non-medical assistance holding income constant across individuals in different health states, optimal insurance could still equate the marginal utility of income across states. In practice, of course, insurance will often fall short of this ideal, but this is nevertheless a useful point of reference. Departures from perfect insurance are discussed further in Section 10.

Thus, expected utility maximization implies that the optimum cost–effectiveness ratio of an intervention in a population under uncertainty is closely approximated by the ratio of expected costs to expected benefits. Note that this “ratio of means” solution is analogous to that suggested by Stinnett and Paltiel (1997) as the solution to a constrained optimization problem in a linear programming context and by Claxton (1999) in a Bayesian discrete choice decision theoretic context. However, neither analysis derives the result directly from a formal utility maximization model nor addresses the possible dependency of the marginal utility of income on θ .

While this argument about the dependence of the marginal utility of income on θ has not been made previously in the context of medical cost–effectiveness analysis, it should be noted that the argument is quite similar to that made by Arrow and Lind (1970) concerning the evaluation of risk in public investment decisions. There the authors argue that the large scale of the public sector allows it to effectively eliminate any welfare loss associated with the riskiness of investments by spreading the risk across a sufficiently large population. The argument here relies both on this diversification effect and the relatively modest magnitude of almost any one public health care decision in the context of overall health and health expenditures.

5. A stochastic population model with individual-level uncertainty about outcomes

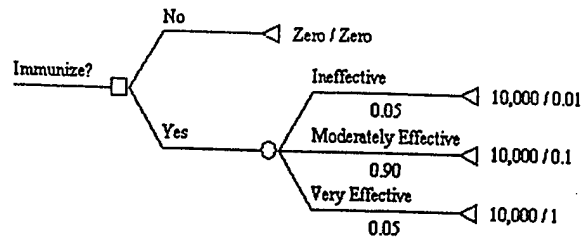
Unlike in the deterministic model presented above, medical interventions almost always have uncertain outcomes for individuals even when there is no population-level heterogeneity so that all individuals share a common set of parameters (θ). Thus, for a set of individuals indexed by $j \in J$ who might each experience health outcome $\varepsilon_j \in E$, the probability of experiencing outcome ε_j given $\theta \in \Theta$ can be written as $f(\varepsilon_j|\theta)$ and expected utility can be written as

$$\int p(\theta) \iint f(\varepsilon_j|\theta) U_j(m, \varepsilon_j, x_j(\varepsilon_j, \theta)) d\varepsilon_j dj d\theta \text{ such that} \\ c_j(m, \varepsilon_j, \theta) + x_j(\varepsilon_j, \theta) - I = 0 \text{ for all } \theta, j, \varepsilon_j. \quad (6)$$

Following the lines of the argument above, we can construct state-specific Lagrange multipliers $\lambda_j(\varepsilon_j, \theta)$ and note that if there is (1) a large population so that aggregate risk given θ is negligible, (2) full insurance, and (3) uncertainty in the effectiveness of the intervention has limited consequences in the sense that θ does not have much effect on λ as described above, then

$$\lim \lambda_j(\bar{\varepsilon}, \theta) \rightarrow \lambda(\bar{\varepsilon}, \theta) \rightarrow \lambda \text{ for all } \bar{\varepsilon} \equiv \{\varepsilon_1, \dots, \varepsilon_j, \dots, \varepsilon_J\}, \text{ where} \\ \varepsilon_j \in E \text{ and for all } \theta. \quad (7)$$

Thus, cost–effectiveness can be identified by the ratio of expected costs to benefits even in the presence of uncertainty at the individual level.



Legend: Cost (\$) / Effectiveness (Life Years)

Fig. 1. Simplified decision concerning a treatment of uncertain benefit (cost (US\$)/effectiveness (life years)).

6. Sensitivity analysis to guide individual or subgroup decisions

When sensitivity analysis is done to guide decisions for individuals or subgroups, the problem is essentially the same as for the total population, except that the parameter vector θ has a different probability distribution $p'(\theta)$ than in the overall population. This may occur if parameters for those individuals or subgroups are thought to differ from those for the population as a whole. This is the type of heterogeneity that most frequently motivates subgroup analyses in cost-effectiveness analysis. However, subgroup analysis may also be desirable if the values of the parameters for a subgroup are not known to differ from those in the population as a whole, but the subpopulation is more or less well studied. In both cases, the analysis differs only in the probability distribution for the parameters, with cases in which some parameters for subgroups are known with certainty addressed by a simplification of the analysis in which the marginal density for the known parameters is degenerate because there is no uncertainty about them.³ Accordingly, the solution to this problem for individuals or subgroups is again the ratio of the expected value of costs to the expected value of benefits, only using the appropriate prior probability distribution for the subgroup or individual.

7. Application to a stylized decision concerning a treatment of uncertain benefit

Fig. 1 describes a stylized decision concerning an intervention of uncertain benefit. For simplicity, the intervention is assumed to cost US\$ 10,000 with certainty. Uncertainty is assumed to exist only with respect to benefits; it is assumed that there is a 90% chance

³ To illustrate: let Groups A and B have pdfs $p^A(\theta)$ and $p^B(\theta)$. Now assume that this heterogeneity can be fully parameterized and partitioned into a certain part (θ_C) and an uncertain part (θ_U) so that these pdfs can be fully parameterized as $p^A(\theta) = p(\theta_U^A; \theta_C^A)$ and $p^B(\theta) = p(\theta_U^B; \theta_C^B)$. In this case, the differences in the certain parameters (θ_C), can be viewed as representing observable heterogeneity, while the uncertainty over the uncertain parameters described by the pdfs describes the uncertainty with respect to which decisions need be made (i.e. integrated over θ_U). Thus, this framework incorporates observable heterogeneity as a special case.

that the benefit is 0.1 life year, but also a 5% chance each that the benefit is 0.01 or 1 life year.

Taking these three possibilities individually, the cost-effectiveness ratios are US\$ 100,000, 1,000,000, or 10,000, respectively. If one used a cutoff of US\$ 100,000 per life year, a traditional sensitivity analysis would therefore be indeterminate. Indeed, such indeterminacy is extremely common in cost-effectiveness analyses. Another limitation of this standard approach is that, while the cost-effectiveness ratios tell us something about the magnitude of benefits relative to costs, they do not provide any indication of how to incorporate the likelihood of those benefits. Common approaches to sensitivity analysis might take other perspectives. For example, the stochastic cost-effectiveness approach might conclude that since there is only a 5% chance that the intervention is not cost-effective, it should be selected. On the other hand, the same approach could be used to argue that since there is only a 5% chance that the intervention will provide a benefit in excess of its cost, it should not be selected. The problem with these perspectives is that they do not reflect the magnitude of potential benefits relative to costs.

Following the expected utility approach described above, the expected cost is US\$ 10,000 and the expected benefit is: $0.05 \times 0.01 + 0.9 \times 0.1 + 0.05 \times 1.0 = 0.0005 + 0.09 + 0.05 = 0.1405$ life years. Thus, the cost-effectiveness ratio is US\$ 10,000 per 0.1405 life years is equivalent to US\$ 71,174 per life year saved, which is clearly cost-effective by the US\$ 100,000 per life year standard. Even though the chance that the intervention is highly beneficial is only 5%, more than one-third ($0.05/0.1405 = 36\%$) of the expected benefit comes from the unlikely event that it is highly effective. It is this ability to incorporate both the magnitude and likelihood of benefits and costs into a single statistic that can be used to guide decision making that is the primary advantage of the expected value approach over the traditional approaches that incorporate only one or the other dimension, and often result in indeterminate conclusions that do not provide much guidance for decision making.

8. Sensitivity analysis to guide information collection

In addition to providing guidance about how to identify the optimal decision under uncertainty given available information, the expected utility approach can be used to inform priorities for research by assessing whether the collection of additional information is likely to be worthwhile. When a study is done to accumulate improved information concerning parameters in a decision model, the value of information is the change in expected utility that comes from a change in uncertainty about the parameters. Although this fundamental principle dates back at least to the pioneers of statistical decision theory (e.g. Raiffa and Schlaifer, 1961; Pratt et al., 1965), and has been used to understand the value of diagnostic testing in medicine (Phelps and Mushlin, 1988), it has not been commonly used to develop techniques for sensitivity analysis in medical decision analysis. Indeed, when formal techniques for clinical trial design have been applied (e.g. O'Brien et al., 1994; Al et al., 1998; Briggs and Gray, 1998), they have often been based on criteria for decision making such as confidence intervals around the cost-effectiveness ratio, which generate suboptimal results for the same reasons that the threshold approaches to sensitivity analysis may be misleading.

Two exceptions to this are Claxton and Posnett (1996) and Hornberger (1998), which focus on the determination of optimal sample size for a clinical trial from a cost-effectiveness perspective in a full Bayesian context.

Adopting the expected utility approach, assume that for any information set describing the parameter distribution, $p(\theta)$, there is an optimal choice of m as described above. Call this $m^*(p(\theta))$. This implies an expected utility with existing information (EU_0) of

$$\int p(\theta) U(m^*(p(\theta)), \theta, x(\theta)) d\theta. \quad (8)$$

Now imagine that we are able to acquire additional information about θ . Assume further that the cost of this research is c_r . Though the analysis is easily generalized to permit an infinite number of possible outcomes of the experiment,⁴ assume for simplicity that there are only two possible outcomes of this experiment: with probability q that the distribution of θ is found to be $p'(\theta)$ and with probability $(1 - q)$ that it is found to be $p''(\theta)$, where, for consistency with the initial prior distribution, $q p'(\theta) + (1 - q) p''(\theta) = p(\theta)$. In these cases, the optimal level of medical expenditure will be $m^*(p'(\theta))$ and $m^*(p''(\theta))$ and the expected level of utility is

$$q \int p'(\theta) U(m^*(p'(\theta)), \theta, x^*(\theta)) d\theta + (1 - q) \int p''(\theta) U(m^*(p''(\theta)), \theta, x^{**}(\theta)) d\theta, \quad (9)$$

where $x^*(\theta)$ and $x^{**}(\theta)$ are determined from the budget constraint net of research costs c_r (i.e. $c(m, \theta) + x(\theta) + c_r - I = 0$ for all θ). It follows that the change in expected utility with the collection of information, or expected value of information (EVI) is

$$q \int p'(\theta) U(m^*(p'(\theta)), \theta, x^*(\theta)) d\theta + (1 - q) \int p''(\theta) U(m^*(p''(\theta)), \theta, x^{**}(\theta)) d\theta - EU_0. \quad (10)$$

⁴ In the general case, we wish to compare the expected utility resulting from the optimal decision m^* given the original budget constraint in the absence of information to the expected utility resulting from the optimal decision in the presence of the new information subject to a budget constraint that includes the cost of collecting information (c_r). Thus, we compare the expected utility resulting from the solution to

$$\max_{m, \lambda(\theta)} \int p(\theta) U(m, \theta, x(\theta)) d\theta + \int \lambda(\theta) p(\theta) [c(m, \theta) + x(\theta) - I] d\theta$$

to the expected utility of the solutions to the optimal decision problems for the j possible outcomes of the experiment as individually given by

$$\max_{m, \lambda(\theta)} \int p_j(\theta) U(m, \theta, x(\theta)) d\theta + \int \lambda(\theta) p_j(\theta) [c(m, \theta) + x(\theta) + c_r - I] d\theta,$$

$$\text{where } \int_{j \in J} p_j(\theta) p(j) dj = p(\theta).$$

Table 1

Information requirements for value of information calculations

Measure of value	Conceptual basis	Information required		
		Burden of illness	Priors for subject of research	Posteriors for subject of research
Expected value of information	Expected gain in welfare from research	Yes	Yes	Yes
Expected value of perfect information	Expected gain from perfectly informative specific experiment	Yes	Yes	
Maximum value of information	Maximum possible gain from specific experiment	Yes	Minimal bounds	
Maximum value of (disease-specific) research	Maximum possible gain for target disease	Yes		

If this is positive then the study is worth performing, if not, then it should not be performed.

Although this value of information calculation is easily described in theoretical terms, implementing this approach requires meaningful information on the prior probabilities of the parameters required for the calculation, and this may be very difficult to obtain. In some instances, priors may be estimated based on published estimates of means and confidence intervals or other data from the literature. In other instances, primary data collection may be required. Still, it is likely that in a significant number of cases it will not be possible to identify much information that will inform priors. Moreover, it may be quite difficult to say much about how an experiment is likely to affect the posterior distributions of the parameters.

These empirical challenges suggest that techniques for assessing the value of information that do not rely on this data concerning prior or posterior distributions would be highly useful. Table 1 summarizes a number of such approaches and their informational requirements. In the case where information on priors is available, one such possibility is the expected value of perfect information: $EVPI = \int p(\theta)U(m^*(\theta))d\theta - EU_0$, where $m^*(\theta)$ is the optimal choice of m if θ is known. Since the expected value of information is always positive,⁵ this provides an upper bound on the ideal value of information calculations described above.

⁵ To see this, note that if research cost are zero, the fact that $m^*(p'(\theta))$ and $m^*(p''(\theta))$ are optima implies that the first two terms in the equation are greater than

$$q \int p'(\theta)U(m^*(p(\theta)), \theta, x(\theta))d\theta + (1-q) \int p''(\theta)U(m^*(p(\theta)), \theta, x(\theta))d\theta \\ = \int p(\theta)U(m^*(p(\theta)), \theta, x(\theta))d\theta. \quad (11)$$

which is the expected utility from the optimal expenditure in the absence of information. This implies that the expected value of free information is positive. For completeness, it should be noted that this result applies only to public information, since the value of private information might not be positive (Rothschild and Stiglitz, 1976).

From a practical point, however, the advantage of the EVPI calculation is that it does not depend on the posteriors. Indeed, this is probably one reason why the EVPI approach has been used in the cost-effectiveness literature (e.g. Felli and Hazen, 1998).⁶

Although EVPI is simpler to determine than EVI, it still depends on knowledge of the priors. An alternative measure that did not depend on this might also be useful. One such measure is the maximal value of information (MVI) over all possible values of $\theta \in \Theta$, $MVB = \max_{\theta \in \Theta} U(m^*(\theta))$. Although this will also only be an upper bound on EVPI and, therefore, EVI, it depends only on knowing the value function conditional on θ . Although it may be a relatively crude upper bound, it is worth noting that this criterion in fact corresponds to that implied by a threshold analysis in which the bounds are determined by the extreme values of the parameter (assuming, as is usually done, that the value function is monotonic with respect to the parameters). Thus, applying the threshold technique based on the full range of possible values of a parameter can be considered a bound on the more general value of information calculation, only with less rigorous information requirements. Thus, like EVPI, the threshold approach based on the full range of values a parameter might take can be considered a method to place an upper bound on the more complex EVI calculation. When these calculations suggest that the MVI or EVPI is low, the full EVI calculation is not necessary. Note, in contrast, that the common practices of assessing cost-effectiveness at a 95% confidence interval for a parameter or calculating stochastic cost-effectiveness intervals have no clear theoretical justification.

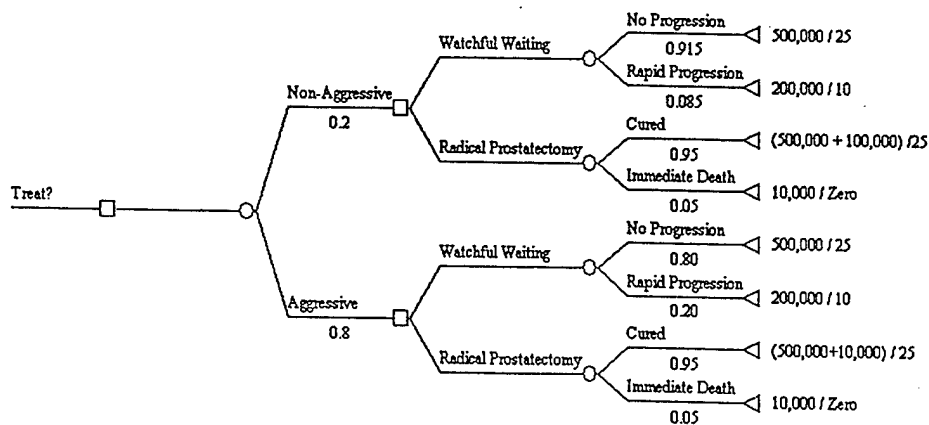
Thinking more broadly, if Θ is enlarged to include any conceivable value of θ , even if the value is not possible with current technology, this type of reasoning can be extended to consider any possible research on the parameter in question. For example, if the probability of cure with the best current treatment for a disease is known to be between 20 and 40% with certainty and the treatment is found not to be worthwhile (perhaps because of morbidity), one could calculate whether treatment would be worthwhile if the cure rate were 100%. This might be called the maximum value of research (MVR), and, in turn, can be used to generate an upper bound on MVI that does not require any data at all concerning the parameter in question. The MVR concept could also be expanded to consider innovations that led to fundamental changes in the structure of the decision tree, and not just the effects of changes in its parameters.

⁶ It should be noted, however, that Felli and Hazen (1998) consider the EVPI relative to the expected value of an optimal decision that they specify as one that maximizes the expected payoff given parameter values that the investigator feels are "most likely to obtain" (p. 100). This seems to suggest the modal value(s) of the parameter(s). Nevertheless, in their applications they tend to choose the mean values of their parameters. Regardless, since neither of these are generally the optimal decision given available information, this calculation will overstate the EVPI relative to the EVPI given an optimal decision with available information. In this sense, Hazen and Felli's calculations can be viewed as an upper bound on the true EVPI. The only advantage of this approach over the theoretically correct approach is that it avoids the need to determine the optimal decision with existing information. This is not a substantial advantage, however, because, if the value of information is small, collecting further information is presumably not worthwhile and therefore knowing the optimal decision with existing information is key. Similarly, if the value of information is large, then one still wants to try to determine the EVPI relative to an optimal decision with existing information to see how much that decreases the bound on the EVPI. Thus, in either case, calculating EVPI relative to the optimal decision given current information is preferred. It is also generally not an extremely difficult determination to make once the ability to assess the expected value of outcomes from an arbitrary (optimal or suboptimal) decision is present.

9. Application to a stylized model of the decision whether to treat prostate cancer

In order to illustrate these approaches, this section examines a simplified model of the decision to treat prostate cancer. A highly stylized model is chosen to focus attention on the methods rather than the specific application. In this simplified model (Fig. 2), the decision to treat prostate cancer is viewed as a choice between radical prostatectomy (surgical removal of the prostate) and “watchful waiting” (no intervention unless the cancer is found to spread). This decision is represented by the two decision nodes in the middle of Fig. 2. In this simplified model, radical prostatectomy is assumed to be curative, so that the patient lives out a “normal” life of 25 years. However, radical prostatectomy is assumed to have a 5% mortality rate. The outcome of watchful waiting depends on how quickly the cancer progresses. Many cancers will progress slowly enough that men die of other causes before they die of prostate cancer and thus live a normal life of 25 years. Other men will progress rapidly and are assumed to die of prostate cancer at 10 years. For simplicity, we assume that quality of life is not a concern so that outcomes are measured in life years, which are the same as quality-adjusted life years. Radical prostatectomy is assumed to cost US\$ 10,000 and the basic future costs of survival are assumed to be US\$ 20,000 per year. (See Meltzer (1997) for a justification for including costs of this nature.)

However, the natural history of prostate cancer is not as well understood as suggested by these assumptions. In fact there is much uncertainty even about average rates of progression to death from prostate cancer, i.e. how aggressive the disease is on average. This is the dimension of uncertainty on which we focus in this example. This is captured in a stylized way in Fig. 2 by the upper and lower decision trees that differ in the fraction of tumors that are assumed to progress rapidly (0.085 in the “non-aggressive” case, and 0.2 in the “aggressive” case).



Legend: Cost (\$) / Effectiveness (Life Years)

Fig. 2. Simplified cost–effectiveness model for screening for prostate cancer with uncertainty about progression rates (cost (US\$)/effectiveness (life years)).

Panels 1 and 2 of Table 1 show the results of a cost-effectiveness analysis of the treatment decision in the non-aggressive and aggressive cases. In both cases, treatment provides a benefit, but in the first case it is a small benefit with a cost per QALY of US\$ 420,000 and in the second case it is a much larger benefit with a cost per QALY of only US\$ 26,000. If we assume for simplicity that the cutoff for cost-effectiveness is US\$ 100,000 per QALY, then the optimal decision in the first case would be watchful waiting, while in the second it would be treatment.

The left most part of the decision tree reflects the fact that we do not know which of these possibilities is the case and places some prior probabilities on the two arms (0.2 aggressive, 0.8 non-aggressive). Panel 3 of Table 1 reports the expected benefits and costs of the screening decision with these priors. In that case, the ratio of the expected costs to expected benefits is US\$ 47,000, which is cost-effective by the US\$ 100,000/QALY standard. This might seem surprising because of the 80% chance that progression was not aggressive, and treatment is not even close to cost-effective by the US\$ 100,000/QALY standard in that case. The result is driven by the 20% chance that the benefit could be much larger, even though that possibility is not very likely. This points out the potential for the ratio of the expected value approach to generate different results than the standard probabilistic approaches based on thresholds for defining cost-effectiveness that do not account fully for both the magnitude and likelihood of the potential benefits.

We now turn to the question of whether the collection of additional information would be of value. Following the approach described above, we begin by calculating the maximum value of information. This calculation can be done in several ways requiring progressively more information. To take an extreme example, assume that we knew nothing about the probability that prostate cancer is aggressive, but only the life expectancy of patients with aggressive cancers who are treated or not treated, and the price of prostatectomy. In the absence of knowledge about the probability that cancers would progress rapidly, there is no clear guidance about whether watchful waiting or prostatectomy dominates, so we consider both cases as reference cases. Assume first that no treatment is the reference point. To get an upper bound on the value of information, one could use only information on the life expectancy of treated and untreated patients and assume that all patients have aggressive cancers. Specifically, assuming that men who have prostate cancer but are not treated live 10 years (QALYs), while men who are treated live 25 years (QALYs), the value of treatment would be $15 \text{ QALYs} \times \text{US\$ } 100,000/\text{QALY} = \text{US\$ } 1.5 \text{ million}$ per patient. Alternatively, we could assume that that treatment is the reference case, so that the benefit of determining that treatment was not cost-effective would be the cost savings from avoiding prostatectomy (US\$ 10,000) and avoidance of treatment-related mortality ($0.05 \text{ mortality} \times 25 \text{ QALYs} \times \text{US\$ } 100,000/\text{QALY} = \text{US\$ } 125,000$) net of any benefits of treatment, which add to no more than US\$ 135,000 per patient.

To use these estimates of the maximum value of information for a patient to assess whether investment in a study to resolve the ambiguity about the aggressiveness of prostate cancer would be worthwhile, one might multiply these numbers by the number of men who are found to have prostate cancer annually (100,000) and divide by some real interest rate (0.03) to reflect the discounted value of the value of that information over time to get the maximum value of information (MVI): $\text{US\$ } 1.5 \text{ million} \times 100,000/0.03 = \text{US\$ } 5 \text{ trillion}$

if the baseline strategy is watchful waiting and $\text{US\$ } 0.135 \text{ million} \times 100,000/0.03 = \text{US\$ } 450 \text{ billion}$ if the baseline strategy is prostatectomy. These extremely large estimates of the maximum value of information suggest the potential for information of immense value to come from knowledge about the efficacy of prostate cancer treatment, and exceed the cost of any conceivable clinical trial.

Of course these MVI calculations are an upper bound, and a fair interpretation of these findings is that the MVI is simply not informative in this case, despite its analytical simplicity and independence of assumptions about the fraction of cancers that are aggressive. This suggests that it is worthwhile to pursue the expected value of perfect information (EVPI) approach.

The EVPI approach is described in panel 4 of the table. The panel describes the expected value of three strategies: watchful waiting, radical prostatectomy, and the optimal decision with perfect knowledge of the average progression rate (EVPI). The last two columns report the value of the change in QALYs (assuming $\text{US\$ } 100,000/\text{QALY}$ for illustration) and the net incremental benefit of the policy choice compared to the strategy immediately above it in Table 2.

The first point to note is that if one made policy based on the most likely cost-effectiveness ratio ($\text{US\$ } 420,000$), one would choose watchful waiting, but if one chose based on the ratio of the expected values, one would choose radical prostatectomy, which yields a net benefit of $\text{US\$ } 19,600$ ($\text{US\$ } 26,000 - 6400$) per patient relative to watchful waiting. This is a quantified measure of the expected gain from the improvement in decision making by using the mean of the expected values as opposed to basing the decision on the most likely cost-effectiveness ratio, as is generally done in the "base case" reported by most current cost-effectiveness analyses.

The second point to note is that the expected value of the gain versus watchful waiting with improved information is even higher at $\text{US\$ } 26,000$ per patient. This implies an additional gain of $\text{US\$ } 6400$ per patient of the improved information compared to the best possible decision with the initial information. Converting this patient level estimate of the value of research into a population level estimate as above suggests an EVPI of $\text{US\$ } 6400 \times 100,000/0.03 = \text{US\$ } 21 \text{ billion}$. As with the MVPI, this large EVPI suggests that the value of information about the efficacy of prostate cancer treatment might far exceed the cost of almost any conceivable clinical trial.

Of course this too is an upper bound on the expected value of information from any actual clinical trial, since any trial is likely to provide less than perfect information. Panel 5 examines one such case in which an experiment has two possible outcomes: a 50% chance of an outcome that suggests that the probability that prostate cancer is aggressive is 0.05 and a 50% chance of an outcome that suggests that prostate cancer is aggressive is 0.35. (Note this preserves the prior that the probability that prostate cancer is aggressive is 0.2 since $0.5 \times 0.05 + 0.5 \times 0.35 = 0.2$.) The expected value of outcomes from watchful waiting and radical prostatectomy given these two possible outcomes of the experiment are reported in the upper and lower parts of panel 5. In the first case, the optimal decision switches to watchful waiting as compared to prostatectomy with the initial information, which yields a net surplus of $\text{US\$ } 600$ per patient. In the second case, prostatectomy remains the optimal choice, so there is no additional benefit to having done the study. Thus, the expected net benefit is $0.5 \times \text{US\$ } 600 = \text{US\$ } 300$ per patient. A decision about the study

Table 2
Value of information for cost-effectiveness of screening for prostate cancer

Strategy	Cost ($\times 1000$ US\$)	Δc ($\times 1000$ US\$)	QALY	$\Delta QALY$	$\Delta c/\Delta QALY$ ($\times 1000$ US\$/QALY)	Value $\Delta QALY$ ($\times 1000$ US\$) ^a	Net increment benefit ($\times 1000$ US\$)
Panel 1. Prostate cancer known non-aggressive; fraction rapidly progressing = 0.085							
Watchful waiting	475	–	23.725	–	–	–	–
Radical prostatectomy	485	11	23.75	0.025	420	2.5	–8.5
Panel 2. Prostate cancer known aggressive; fraction rapidly progressing = 0.2							
Watchful waiting	440	–	22	–	–	–	–
Radical prostatectomy	485	45	23.75	1.75	26	175	130
Panel 3. Aggressiveness of prostate cancer not known; probability aggressive (as in panel 2) = 0.2							
Watchful waiting	467.6	–	23.38	–	–	–	–
Radical prostatectomy	485	17.4	23.75	0.37	47	37	19.6
Panel 4. Expected value with perfect information: probability aggressive (as in panel 2) = 0.2							
Watchful waiting	467.6	–	23.38	–	–	–	–
Optimal w/perfect information	476.6	9	23.73	0.35	26	35	26
Radical prostatectomy	485	8.4	23.75	0.02	420	2	–6.4
Panel 5. Expected value w/improved information: 50% chance study \Rightarrow probability aggressive = 0.05; 50% chance study \Rightarrow probability aggressive = 0.35 ^b							
Probability aggressive = 0.05							
Watchful waiting	473.3	–	23.64	–	–	–	–
Radical prostatectomy	485	11.7	23.75	0.11	106	11.1	–0.6
Probability aggressive = 0.35							
Watchful waiting	462.8	–	23.12	–	–	–	–
Radical prostatectomy	485	22.2	23.75	0.63	35	62.9	40.6

^a 1 QALY = US\$ 100,000.

^b Expected value of information (vs. expected value of optimal decision with initial information (prostatectomy)) = $0.5 \times 0.6 = 0.3$.

might be made by comparing its cost to the expected value of the information (EVI): $\text{US\$ } 300 \times 100,000/0.03 = \text{US\$ } 1 \text{ billion}$. Therefore, the value of this study would be quite large, although substantially less than the upper bound suggested by the EVPI.

In a similar manner, possible experiments concerning all other dimensions of the model might be examined to determine whether they would be worthwhile. In this way, it might be determined how much could be gained by improved sensitivity and specificity of screening tests, decreased complications of treatment, improved risk stratification prior to treatment, and so on.

Clearly, this example does not suggest that a comprehensive attempt to perform a precise calculation of the type described would generate results resembling these in magnitude. However, these simplified calculations do illustrate the types of calculations that might be used to assess the value of research, including more simple calculations such as the EVPI that require less information. The results also suggest, however, the potential for some of the approaches used in the literature, such as the threshold (MVI) or EVPI to provide only very crude upper bounds on the value of information. Just how informative such bounds may be in practice will ultimately be determined only by detailed empirical analysis of specific clinical applications.

10. Conclusion

This paper has examined the purposes for which sensitivity analysis is performed in medical cost–effectiveness analysis and the implications of an expected utility maximization model for the methods to perform such analyses. The analysis suggests specific approaches for optimal decision making under uncertainty, specifying such decisions for subgroups, and assessing the value of collecting additional information.

At a theoretical level, there are several limitations of this work. First, even with certainty about costs and benefits, cost–effectiveness analysis may not maximize the welfare of individuals (Meltzer et al., 1998), or society (Arrow, 1951; Meltzer and Johannesson, 1998). Perhaps more important are issues about how risk at the individual level may affect welfare (Kahneman and Tversky, 1979) that are essentially ignored by the assumptions of perfect insurance and expected utility maximization. Though this is an important limitation of QALYs, it is one that needs to be addressed regardless of the issues about aggregate uncertainty addressed by sensitivity analysis. Though concerns about aggregate financial and health risk may be less compelling in a social context where the aggregate risks associated with individual technologies are usually modest, the issue of how risk should be assessed in policy decisions deserves further consideration because other assumptions about preferences concerning risk or about insurance would lead to different conclusions about many methodological issues in cost–effectiveness analysis, including sensitivity analysis (e.g. Mullahy, 1997). Indeed, when a medical intervention has major financial implications that are difficult to insure against, such as lost earnings, the marginal utility of income cannot reasonably be considered constant and the results above concerning the ratio of means will no longer hold. This suggests that it may be useful to distinguish between uncertainty in insured and uninsured costs in assessing the implications of uncertainty in costs in cost–effectiveness analyses. Additionally, it suggests that further characterization of

optimal decision making when insurance is not complete would be a valuable area for future work.

Rather than using expected utility to incorporate preferences over uncertain outcomes, it might be argued that it would be preferable to report the joint distribution of benefits and costs. Nothing about this analysis suggests that such data should not be presented. However, using such data to make choices would still require decisions about how to incorporate risk into decision making. Unlike traditional forms of sensitivity analysis, the expected value approach provides direct guidance about how the optimal decision varies with the assumptions that are made.

At an empirical level, there are important challenges in developing meaningful priors concerning the parameters of decision models (e.g. probabilities, quality of life values, discount rates, etc.). As discussed above, this may often require extensive review of existing data, primary data collection, or even analyses based on arbitrary priors. It may also be very difficult to specify how research may affect posteriors. Whether it is possible to adequately address these challenges will be resolved only through efforts to apply these ideas empirically.

These approaches to assess the value of research also pose additional challenges. These include the interdependence of the benefits of related research, the possibility that the research might become less (or more) valuable over time if technological or demographic changes alter the management, frequency or natural history of a disease, and the unpredictability of how the results of research (particularly basic research) might be useful in areas outside the initial areas of inquiry (serendipity). The difficulty of these issues implies that the sort of formal analyses suggested here are more likely to be useful for evaluating clinical research than basic research.

Despite these theoretical and empirical challenges, the importance of making good decisions about the allocation of resources to medical interventions and medical research suggest that work in this area be an important priority. It is encouraging in this regard that the recent IOM report on improving priority setting at the NIH recommended: "In setting priorities, NIH should strengthen its analysis and use of health data, such as burdens and costs of diseases, and on data on the impact of research on the health of the public" (IOM, 1998, p. 11).

On the other hand, the limited number of cases where cost-effectiveness analysis has strongly influenced medical resource allocation and the likely resistance of medical researchers to having research proposals evaluated by formal criteria suggest that formal techniques to set priorities for research will have to prove their value. It is possible that cost-effectiveness analysis may enhance its influence if it can address key methodological challenges in measuring benefits and costs, and techniques for sensitivity analysis. There may also be less resistance to the use of cost-effectiveness analysis in policy decisions, such as allocation of research funds, than to its use in decisions to ration medical treatments. Nevertheless, formal techniques to inform priorities for research seems more likely to gain acceptance through instances where neglected areas of research can be identified through formal analysis than through instances where research is suggested to be of little value. Consistent with this, threats to increases in the NIH budget due to Congressional questions about the value of increased appropriations for research and NIH priorities in allocating research funds were an important motivation for the IOM report that encouraged efforts to use formal approaches to determine the value of research.

Acknowledgements

The author gratefully acknowledges the financial support of this work by the National Institute of Aging, the Robert Wood Johnson Generalist Physician Faculty Scholars Program, and the Department of Defense Prostate Cancer Research Project. I would also like to thank Joshua Angrist, David Cutler, Sue Goldie, Zvi Griliches, Willard Manning, John Mullahy, and Milton Weinstein for helpful comments on earlier drafts of this paper.

References

- Al, M.J., van Hout, B.A., Michel, B.C., Rutten, F.F.H., 1998. Sample size calculation in economic evaluations. *Health Economics* 7, 327–335.
- Arrow, K., 1951. *Social Choice and Individual Values*. Yale University Press, New Haven, CT.
- Arrow, K., Lind, R., 1970. Uncertainty and the evaluation of public investment decisions. *American Economic Review* 16 (3), 364–378.
- Briggs, A.H., Gray, A.M., 1998. Power and significance calculations for stochastic cost–effectiveness analysis. *Medical Decision Making* 18 (Suppl.), S81–S92.
- Brook, R., et al., 1983. Does free care improve adults' health? Results from a randomized controlled trial. *New England Journal of Medicine* 309 (24), 26–1434.
- Claxton, K., 1999. The irrelevance of inference: a decision making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics* 18 (3) 341–364.
- Claxton, K., Posnett, J., 1996. An economic approach to clinical trial design and research priority-setting. *Health Economics* 5 (6) 513–524.
- Cutler, D., Richardson, E., 1997. Measuring the health of the US population. *Brookings Papers. Microeconomics* 217–271.
- Eddy, D., 1990. Screening for cervical cancer. *Ann. Int. Med.* 113, 214–226.
- Elixhauser, A., Halpern, M., Schmier, J., Luce, B.R., 1998. Health care CBA and CEA from 1991 to 1996: an updated bibliography. *Medical Care* 36 (5), MS1–MS9.
- Felli, J.C., Hazen, G.B., 1998. Sensitivity analysis and the expected value of perfect information. *Medical Decision Making* 18 (1), 95–109.
- Fuchs, V., 1990. The health sector's share of the gross national product. *Science* 247, 534–538.
- Gold, M.R., Siegel, J.E., Russel, L.B., Weinstein, M.C., 1996. *Cost-Effectiveness in Health and Medicine*. Oxford University Press, New York.
- Hornberger, J., 1998. A cost–benefit analysis of a cardiovascular disease prevention trial using folate supplementation as an example. *American Journal of Public Health* 88 (1), 61–67.
- Institute of Medicine, 1998. *Scientific Opportunities and Public Needs: Improving Priority Setting and Public Input at the National Institutes of Health*. National Academy Press, Washington, DC.
- Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47 (2), 263–291.
- McClellan, M., McNeil, B.J., Newhouse, J.P., 1994. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? *Journal of American Medical Association* 272 (11), 859–866.
- Meltzer, D., 1997. Accounting for future costs in medical cost–effectiveness analysis. *Journal of Health Economics* 16 (1), 33–64.
- Meltzer, D., Johannesson, M., 1998. Inconsistencies in the 'societal perspective' on costs of the panel on cost–effectiveness in health and medicine. Unpublished manuscript.
- Meltzer, D., Polonsky, T., Egelston, B., Tobian, J., 1998. Does quality-adjusted life expectancy predict patient preferences for treatment in IDDM: validating quality of life measures through revealed preference. Unpublished manuscript.
- Mullahy, J., 1997. Which cost–effectiveness ratio? Evaluating health policies and medical technologies in stochastic contexts. University of Wisconsin, Madison, Unpublished manuscript.

- Murphy, K.M., Topel, R.H., 1998. The economic value of medical research. University of Chicago, MD, Unpublished manuscript.
- Newhouse, J.P., 1992. Medical care costs: how much welfare loss? *Journal of Economic Perspectives* 6, 3–21.
- O'Brien, B.J., Drummond, M.F., Labelle, R.J., Willian, A., 1994. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Medical Care* 32 (2), 150–163.
- Phelps, C.E., Mushlin, A.I., 1988. Focusing technology assessment using medical decision theory. *Medical Decision Making* 8 (4), 279–289.
- Polsky, D., Glick, H.A., Willke, R., Shulman, K., 1997. Confidence intervals for cost-effectiveness analysis: a comparison of four methods. *Health Economics* 6, 243–252.
- Pratt, J.W., Raiffa, H., Schlaifer, R., 1965. *Introduction to Statistical Decision Theory*. McGraw-Hill, New York.
- Raiffa, H., Schlaifer, R., 1961. *Applied Statistical Decision Theory*. Harvard Business School, Colonial Press.
- Rothschild, M., Stiglitz, J., 1976. Equilibrium in competitive insurance markets: an essay on the economics of imperfect information. *Quarterly Journal of Economics* 90, 629–650.
- Smith, S., et al., 1998. The next ten years of health spending: what does the future hold? *Health Affairs* 17 (5), 128–140.
- Stinnett, A.A., Mullahy, J., 1998. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Medical Decision Making* 18 (Suppl. 2), S68–S80.
- Stinnett, A.A., Paltiel, D., 1997. Estimating CE ratios under second-order uncertainty: the mean ratio versus the ratio of the means. *Medical Decision Making* 17, 483–489.
- Tambour, M., Zethraeus, N., Johannesson, M., 1998. *International Journal of Technology Assessment in Health Care* 14 (3), 467–471.

Appendix 6:

**Inconsistencies in the "Societal Perspective" on Costs of the Panel on Cost-
Effectiveness in Health and Medicine**

David Meltzer , Magnus Johannesson

Medical Decision Making

19:371-377,1999

Forum: Recommendations for Cost-Effectiveness Analysis

Inconsistencies in the "Societal Perspective" on Costs of the Panel on Cost-Effectiveness in Health and Medicine

DAVID MELTZER, MD, PhD, MAGNUS JOHANNESSON, PhD

A key recommendation of the recent Panel on Cost-Effectiveness in Health and Medicine was that cost-effectiveness analyses be carried out from a societal perspective. The authors show that two of the Panel's recommendations concerning costs are not consistent with a societal perspective, and how to correct those inconsistencies. In its recommendations concerning costs resulting from morbidity, the Panel advises excluding lost income from costs in the belief that individuals take income changes into account when they respond to the quality-of-life questions that are used to calculate quality-adjusted life years (QALYs). It is shown that even if individuals do consider income changes in responding to these quality-of-life questions, this recommendation would seriously underestimate production losses due to morbidity, since individuals do not bear a major part of lost production. In its recommendations concerning costs resulting from mortality, the Panel does not require that health care costs for "unrelated" illness and non-health care consumption and production during added life years be included in the Reference Case. It is shown that omitting these costs will seriously distort comparisons of programs at different ages and favor programs that extend life over those that improve quality of life. This can be corrected by including total consumption minus production in added life-years among costs. *Key words:* cost-effectiveness analysis; societal perspective; public policy; resource allocation; Panel on Cost-Effectiveness in Health and Medicine. (*Med Decis Making* 1999;19:371-377)

The rapid advancement of medical technology and the resulting increase in health care costs have increased interest in the economic evaluation of health care programs. There are various methods for the economic evaluation of medical interventions, but the one that is currently most popular is cost-effectiveness analysis. Following this approach, costs are measured in monetary terms and effectiveness is measured in non-monetary terms, e.g., quality-adjusted life years (QALYs). A problem with the method is that it has not been clear how to de-

fine costs and effectiveness. This has led to great variability in what is included in measures of costs and effectiveness, which has caused problems in comparing and interpreting the results of cost-effectiveness analyses.¹

Because of these problems, the Panel on Cost-Effectiveness in Health and Medicine was convened in 1993 by the U.S. Public Health Service to develop recommendations for the conduct of cost-effectiveness analysis. The official report of the Panel² was summarized in a series of three papers in JAMA.³⁻⁵ A key recommendation of the Panel was that cost-effectiveness analyses be carried out from a societal perspective, including all the costs and consequences of health interventions no matter to whom they accrue. However, the Panel also recommended that most productivity gains and losses resulting from medical interventions that reduce morbidity (often referred to as morbidity costs) not be included among the costs in cost-effectiveness analyses. In addition, they concluded that it was not necessary to include certain classes of medical or non-medical expenditures in additional years of life

Received October 6, 1997, from the Section of General Internal Medicine, Department of Economics, and Harris Graduate School of Public Policy Studies, University of Chicago, Chicago, Illinois (DM); and the Center for Health Economics, Stockholm School of Economics, Stockholm, Sweden (MJ). Revision accepted for publication April 21, 1999. Supported in part by a grant from the National Institute of Aging and The Robert Wood Johnson Foundation (Meltzer).

Address correspondence and reprint requests to Dr. Meltzer: Section of General Internal Medicine, University of Chicago, 5841 South Maryland MC 2007, Chicago, IL 60637; phone: (773) 702-0836; fax: (773) 834-2238; e-mail: <dmeltzer@medicine.bsd.uchicago.edu>.

that result when a medical intervention reduces mortality (often referred to as mortality costs).

Here we argue that these recommendations of the Panel are not fully consistent with a societal perspective since they do not correctly account for costs resulting from the effects of changes in morbidity on productivity and the effects of mortality on productivity and consumption in added years of life. This implies that cost-effectiveness analyses carried out in accordance with the recommendations of the Panel would not lead to an efficient use of resources from a societal perspective. Below we show why the recommendations of the Panel are inconsistent with a societal perspective, and how the definition of costs in cost-effectiveness analysis can be made consistent with a societal perspective.

The Treatment of Morbidity Costs

Apart from affecting health care costs (and other direct costs), a health care program may also affect the production of goods and services by an individual, whether in the market or outside the market. For simplicity, we can call production inside and production outside the market, respectively, "income" and "leisure." These effects of morbidity on production are recognized by the Panel. However, the Panel argues that changes in market and non-market production borne by individuals should not be included among the costs.^{2,p.306} Their justification for this recommendation is based on the idea that people already include these costs in estimating quality-of-life weights used to construct QALYs, so that it would be double-counting to include them among the costs. The Panel also recognized that not all productivity losses would be borne by individuals and concluded that "effects of lost productivity borne by others (e.g. employers, coworkers), including 'friction costs,' when significant, should be included in the numerator."^{2,p.186}

While the general principle described by the Panel that effects not incorporated in the denominator be included in the numerator is theoretically sound, it is not clear that their assumptions concerning what people actually include when responding to the QALY questions used to construct the denominator are correct. In particular, it is not known whether individuals take changes in income or leisure into account when they respond to the quality-of-life questions used to construct QALY weights. It seems plausible that changes in leisure time may be incorporated in the QALY weights, since individuals may weigh these effects as they respond to questions evaluating quality of life in specific health states. However, the argument that changes in market production are reflected in the quality weights is not credible. One instrument used to generate QALY

weights for health states, the Health Utilities Index, explicitly instructs individuals being interviewed to assume that their financial circumstances would not vary with health status.⁶ All other elicitation processes provide respondents with no guidance about what to assume about the economic consequences of health states when responding to QALY elicitation questions. Without guidance, it is possible that individuals may or may not incorporate financial effects.

The Panel attempts to address this concern by recommending that "for the reference case, health-related quality of life should be captured by an instrument that, at minimum, implicitly incorporates the effects of morbidity on productivity and leisure."^{2,p.306} Thus, the Panel's recommendations imply that instruments such as the Health Utilities Index should not be used for utility assessment. However, the Panel also recommends that, "when instruments used to measure health states are silent concerning the consideration of lost income, we assume that financial effects have been considered by the respondent and that it is therefore not necessary to account for these effects in the numerator."^{2,p.209,4} Unfortunately, there is no evidence that this assumption is valid. To have confidence in the Panel's position that personal financial effects are already incorporated in QALY weights, either respondents should be explicitly told to include these effects or empirical research should be performed to demonstrate that this is the case.

Moreover, even if individuals do consider personal financial consequences in responding to QALY elicitation, the recommendations of the Panel are not consistent with a societal perspective because they do not adequately reflect the extent to which individuals would undervalue changes in production if they were to take only personal financial consequences into account in responding to QALY questions. In particular, although the Panel did recognize the need to include costs borne by others such as employers and coworkers,^{2,p.306} they did not adequately recognize the variety of forms of insurance and taxes that prevent individuals from bearing the full consequences of changes in their production. When there is fully paid public or private disability insurance, for instance, a person does not experience any change in income when he or she experiences illness, and might therefore not consider effects on productivity when responding to QALY questions. Nevertheless, the Panel advises excluding payments such as these from costs on the grounds that they are transfer payments.^{2,p.307} The result is that the real decrease in productivity resulting from the illness is reflected neither in the numerator nor in the denominator, so that the analysis is inconsistent with a social perspective.⁷ Al-

though transfer payments do not represent true social costs, the accounting approach advocated by the Panel results in the rather unintuitive implication that the true social costs of lost productivity would be captured only by including these transfer payments from disability insurance as a cost.

But even in the absence of public or private disability insurance, the presence of income and payroll taxes implies that an individual bears only a fraction of the consequences of changes in production. This is because the true value of the lost production of a worker who is absent from work is equal to the amount of money the employer is willing to pay for that work, which also includes the income tax paid by the worker and the payroll tax paid by the employer. It is important to note that, unlike productivity losses related to "friction costs," which have been the focus of several recent articles discussing the appropriate treatment of productivity costs,⁷⁻⁹ and rest on debatable assumptions about the rigidity of labor markets, these costs last as long as the disability lasts and may be very large. To illustrate the magnitude of this effect, note that with a payroll tax of 30% of the wage rate and an income tax of 30%, the individual will bear only about 50% of the value of the change in production due to disease even if no payment for sick leave exists. Failing to account for these changes in production will thus seriously underestimate the value of gains or losses in production, even if individuals reflect the changes in personal income in responding to questions that elicit QALY weights. Unfortunately, the Panel never discusses the question of whether income and payroll taxes should be included in costs except in one table, where the implication appears to be that they should not be included.^{2,p.187}

The Panel also advised including all health care costs in measures of costs. The problem with this recommendation is that if people incorporate personal financial consequences into QALY weights, they should presumably also incorporate out-of-pocket medical expenditures, and including all health care costs in measures of costs would lead to double-counting of the out-of-pocket costs. In the United States, this could lead to substantial errors in estimates of both the absolute and the relative cost-effectiveness of medical interventions, since about 20% of health care expenditures are paid for out-of-pocket, and the fractions paid out-of-pocket vary a great deal across different types of medical procedures, depending on insurance coverage.¹⁰

A similar problem arises with the Panel's recommendation to include the value of time lost from work or leisure activities in order to participate in a health intervention (such as an exercise program) in measures of costs. If individuals incorporate effects

on personal economic status and leisure into QALY weights, they should presumably already be incorporating these effects into their QALY weights, and including them in measures of costs as well would again lead to double-counting of these costs. This may not be much of a problem when the treatment is infrequently required, but could be very important when the treatment is time-consuming and chronic, as is hemodialysis for end-stage renal disease.

Therefore, even if individuals consider the private economic consequences of illness in responding to QALY elicitation, the recommendations of the panel are not consistent with a societal perspective on costs. Indeed, it is only under the very specific assumption that individuals incorporate into their QALY weights exactly what the Panel has chosen to exclude from costs that the Panel's recommendations can be considered to be consistent with a societal perspective on costs. To be precise, this means that individuals' QALY weights would have to reflect the value of lost leisure and the total value of lost productivity in the market, including not only take-home wages but also taxes that the government would have received from individuals as income taxes and from their employers as payroll taxes. Moreover, individuals would have to incorporate these losses into their QALY weights only if the productivity change was due to illness and not if it was due to participation in a health care intervention. Also, despite reflecting these changes in their personal welfare due to changes in income and leisure in their QALY weights, individuals would have to exclude changes in medical costs from their QALY weights, even if the costs were paid out of pocket.

Thus, the Panel's assumption that, when health status measures are silent on financial issues, financial effects have "been considered by the respondent"⁴ in such a way as to lead to consistent definitions of benefits and costs seems extremely unlikely to be valid. Moreover, even the recommendation that respondents be explicitly told to "consider the full range of impacts of the health status change, including loss of income and leisure activities"⁴ will not generate responses consistent with a societal perspective on costs. Indeed, the Report² recognizes that when people do not take "effects of lost productivity borne by others" into account [p. 306], those effects must be included in costs. Nevertheless, the report does not follow through with that principle by recommending that lost tax revenue be included in costs or that out-of-pocket expenditures on health care be excluded from health care costs.

These inconsistencies in the Panel's recommendations could be corrected by clarifying the process of QALY elicitation to be explicit about what eco-

nomic consequences individuals are to consider and then using an appropriate definition of costs. One possibility, which is similar to the recommendations of the Panel, would be to encourage individuals to take into account all the personal financial consequences of changes in their market and non-market productivity due to illness (i.e., changes in after-tax income and out-of-pocket medical costs and changes in leisure) when responding to QALY elicitation questions. However, in that case, the part of health care costs and production change not borne by the individual should be included in the estimation of costs. Alternatively, individuals could be told to provide QALY weights based on the assumption that their financial circumstances would not vary with health status, but that their non-market productivity would vary with health status (i.e., that they have paid sick leave and do not pay anything out of pocket for health care costs).^{*} The full health care costs and the full change in production should then be included in the cost estimation.

Either of these approaches could be justified in theory, but from a practical viewpoint the second approach seems preferable, since the analysis becomes much more straightforward to perform and to interpret. It avoids collecting detailed information about paid sick leave and out-of-pocket payments, which would be necessary following the first approach. The advantage of the second approach is even more obvious if QALY weights, as recommended by the Panel, are elicited from the general population without experience of the health states to be assessed. This makes the second approach preferable, because otherwise individuals would have to make uninformed guesses about the incomes in different health states, or the income level would have to be explicitly incorporated as an attribute in the health status classification system used.

The Treatment of Mortality Costs

The correct treatment of mortality costs is a long-standing and active area of controversy in medical cost-effectiveness analysis. As recognized by the Panel, and further clarified subsequent to the publication of its report, the theoretically correct way to include mortality costs is to include the difference

^{*}It may require some experimentation to determine exactly the best approach to doing this, but we would propose the following sentence as one concrete possibility. For example, in introducing a time-trade-off question, we might suggest the statement: "In answering this question, please assume that your financial circumstances would not be affected by the illness, as might be the case if you did not have to pay anything for health care and you received disability payments that completely replaced any lost earnings."

between total consumption and total production during life years gained by a medical intervention as a cost in cost-effectiveness analysis.¹¹⁻¹³ However, the Panel did not recommend that these costs be included in all cases. In the case of health care costs for "unrelated" illnesses during life years gained by an intervention, they concluded that cost-effectiveness ratios can be estimated with or without including these costs, but that a sensitivity analysis should be performed whenever those costs had a significant effect on the results. In the case of non-health care costs during gained life years, the summary recommendations of the Panel did not include any relevant recommendation, though other parts of the report suggested conditions under which these costs could be omitted^{2, pp. 48, 186} and a later article summarizing its recommendations suggests that these costs be excluded.⁴

The Panel discusses several different arguments to justify excluding future costs [p. 186]. One is purely political—namely that people may not be willing to accept including such costs in medical cost-effectiveness. We postpone discussion of this argument until the conclusion. The other arguments are summarized as follows:

In addition to this difference of opinion, there are practical difficulties in including costs for unrelated illness in added years of life. Existing data may not be adequate to capture future resource use of all unrelated diseases; in addition, it may be unduly difficult to ascertain the effect of an intervention on the range of future causes of morbidity and death. Finally, if these costs are included, non-health care costs in added years of life should also be included. As discussed in Chapter 2, the exclusion of non-health care costs is acceptable if these costs add a constant cost to each year; however, no research has been done to determine whether this is the case.

Because of the practical concerns and unresolved theoretical issues surrounding the inclusion of future costs for unrelated illness in added years of life, our recommendation for the Reference Case is that analysts use their discretion in including or excluding these costs. Like other costs and consequences, the rule of reason applies to these costs: If they are small compared to the magnitude of the C/E ratio, they can be omitted without affecting the analysis results in any case. If they are large, we recommend that the analyst conduct a sensitivity analysis to assess their effect." [p. 186]

To assess the Panel's position on future costs, it is useful to review these arguments systematically. First, there is the issue of the adequacy of data to estimate future costs for unrelated diseases. While the Panel is correct that this may be difficult to do precisely and is an area where further research

would be valuable, they also describe a number of data sources with which "first" estimates could be derived [pp. 186-8] and there are now estimates in the literature.^{12,14} In any case, it seems difficult to argue that including an implicit estimate of zero by omitting unrelated costs would be preferable to an imprecise estimate, especially with appropriate sensitivity analysis. It is not clear what the Panel means by difficulty in ascertaining "the effect of an intervention on the range of future causes of morbidity and death," but if it concerns related illnesses, then the problem applies equally to costs for related illness, whose inclusion is recommended by the Panel. If it concerns unrelated illness, the very definition of such costs suggests that they should be unaffected by the intervention and, again, it seems unlikely that an estimate of zero would be preferable to an imprecise estimate with appropriate sensitivity analysis.

The recommendation that whenever future costs for unrelated illness are included, future non-health care costs also be included is consistent with theory, but implies that it is acceptable not to include such non-health care costs so long as costs of unrelated illnesses are not included. As we show below, this is not only inconsistent in theory, but also causes a significant bias of cost-effectiveness analyses in practice.

The statement that the exclusion of non-health care costs is acceptable if these costs add a constant cost to each year relates to a theoretical finding by Garber and Phelps¹⁵ that if future costs are conditionally independent in the sense that, conditional on a person's surviving, the costs do not vary with the medical intervention in question, then including those costs will only add a constant to the cost-effectiveness ratio. Even if conditional independence holds, it is important to note (as does the Panel itself^{2,p.46}) that this result holds only when comparing interventions at a single age, since future costs due to a decrease in mortality will vary widely depending on age.¹² For example, a recent study of hypertension treatment demonstrated that the correct inclusion of these costs would not significantly affect the cost per QALY gained for younger men and women, but would increase the cost per QALY gained by nearly \$30,000 for older men and women.¹⁴

Moreover, even when comparing interventions at a single age and when future costs are "conditionally independent," the argument turns out to be incorrect. A simple example can illustrate this. Imagine we wish to compare the cost-effectiveness of two completely independent interventions (A and B) for two completely separate but both otherwise fatal diseases at a single age. Assume that each generates one QALY, but that A adds one year of life at QOL

= 1 and B adds two years of life at QOL = 0.5. Thus intervention A has bigger effects on quality of life than on length of life, while intervention B has bigger effects on length of life than on quality of life. Assume that the immediate related costs are \$5,000 for A and \$1,000 for B. Let annual related costs be zero and annual unrelated costs (of consumption minus production) be \$20,000 for both. Thus, the definition of conditional independence is met. Excluding future unrelated costs, $C/E(A) = \$5,000/\text{QALY}$ and $C/E(B) = \$1,000/\text{QALY}$. Including future costs, the cost-effectiveness ratios are $(\$5,000 + \$20,000)/1 \text{ QALY} = \$25,000/\text{QALY}$ for A and $(\$1,000 + 2 \cdot \$20,000)/1 \text{ QALY} = \$41,000/\text{QALY}$ for B. Therefore, including future costs changes both the absolute and the relative cost-effectiveness of the two interventions even when the strict definition of conditional independence is met. Thus, in contradiction to the findings of the Panel, the correct inclusion of future cost is not an empirical question depending on whether future costs add a constant to costs in each year. This example also illustrates the more general point that excluding future costs biases cost-effectiveness analyses of interventions among groups with positive net future resource use (such as the elderly) to favor interventions that extend life over interventions that improve the quality of life.¹² Conversely, excluding future costs among groups with negative future net resource use (such as the young) will favor interventions that improve quality of life over those that increase length of life. Estimates of the magnitude of these effects of including future costs suggests that they are large enough to significantly alter the relative cost-effectiveness of interventions such as the treatment of hypertension, chemotherapy for colon cancer, and hemodialysis for end-stage renal disease.¹²

Therefore, the recommendation in the second paragraph in the section of the report cited above that suggests that analysts can use their discretion about the inclusion of future unrelated costs in the Reference Case is not consistent with either theoretical concerns or strongly defensible practical considerations. It also leaves the literature open to being saddled with one set of analyses that do not include future costs and are not comparable to those that do include them. The recommendation that a sensitivity analysis be conducted if these effects are large begs the question, since it does not describe how one is to know whether future costs are important without performing such an analysis. Likewise, the rule of reason argument begs the question, since it does not define what constitutes a significant difference.

For these reasons, both theoretical and empirical, it seems to us most appropriate to include future costs in the Reference Case. If a sensitivity analysis

suggests that their effect is trivial, as might be expected when there is no or almost no effect on survival, then omitting these costs may not be problematic. However, an adequate sensitivity analysis should always precede the decision to omit such costs, and the decision to omit those costs should be made only when the resulting change in the cost-effectiveness ratio is so small as to be insignificant for any sort of statements of relative or absolute cost-effectiveness. How small is small here depends on the importance one places on small differences in cost-effectiveness ratios.

Conclusion

We fully agree with the Panel that cost-effectiveness analyses should be carried out from a societal perspective, including all costs and benefits no matter to whom they accrue. Furthermore, we applaud their efforts to move cost-effectiveness analysis closer towards that important objective.

Unfortunately, several of the recommendations of the Panel concerning costs are not fully consistent with a societal perspective. In the case of morbidity costs, the Panel's recommendation to exclude changes in production from costs is likely to lead to underestimating the cost-effectiveness of interventions that enhance productivity by decreasing morbidity. For the reasons discussed above, we believe this would be best addressed by instructing individuals to provide QALY weights based on the assumption that their financial circumstances would not vary with health status (i.e., that they have paid sick leave and do not pay anything out of pocket for health care costs), but that their non-market productivity would vary with health status. Further empirical work is required to determine what approaches to quality-of-life assessment and cost measurement can best address this problem, and we feel strongly that such work should have a high priority.

Similarly, the Panel's recommendations concerning mortality costs will substantially distort the comparison of cost-effectiveness across different ages and tend to inappropriately favor programs that extend life over programs that improve the quality of life. To use a societal perspective for cost-effectiveness analysis, it is necessary to include consumption minus production during life years gained as a cost in the analysis. While including such costs may sometimes not make a large difference, there is good evidence that they may also commonly make a substantial difference. We therefore propose that a sensitivity analysis be done to assess the magnitude of future costs before a decision is made to omit them in a particular analysis.

One issue we have neglected to this point is the

political acceptability of including future costs. Implicitly, there may be a feeling that this will result in denial of care to the elderly, and that this is somehow "unfair." We are sympathetic to these concerns, but return to the fact that these costs are real, and note that the vast majority of cost-effective interventions among the elderly will continue to be cost-effective when future costs are included, and that failing to account for future costs biases cost-effectiveness analyses to favor interventions that extend life over interventions that improve the quality of life. If QALYs do not result in a distributional outcome that is desirable to society, we propose creating models that are explicitly based on those values and then performing analyses to identify interventions that do help realize them. If there are costs or benefits we have failed to model in cost-effectiveness analyses, we favor modeling them explicitly as well. The alternative approach of ignoring costs we know to be real tends only to create additional biases, such as favoring length of life over quality of life among the elderly.

The work of the Panel on cost-effectiveness in Health and Medicine is a useful step in the methodologic development of medical cost-effectiveness analysis. Nevertheless, the presence of these important inconsistencies in the Panel's recommendations concerning costs demonstrates the need to continue to evaluate the methods of cost-effectiveness analysis if the field is to realize its promise to improve the allocation of scarce resources for health care. As the Report of the Panel effectively argues, this will surely require attention to both the theoretical foundations of cost-effectiveness analysis and the simplicity, feasibility, and broad acceptance of its methods. It is possible that some of the Panel's conclusions on these issues may have been shaped by the latter factors, in which case it is our hope that these issues may come to be reconsidered over time as methodologic advances in areas such as quality-of-life assessment and estimation of future costs make us more able to address these issues successfully. Indeed, the need for ongoing, critical appraisal of the methods of cost-effectiveness analysis is reinforced by the fact that cost-effectiveness analysis is by no means the only approach to allocation of resources within health care. Though they also have their own limitations, cost-benefit analyses that assess willingness to pay, formal and informal reflections of expert and public opinion, and greater attention to the development of the knowledge and incentive systems to help patients and their physicians make better decisions concerning medical spending are all alternatives worthy of consideration. The presence of these important alternatives to cost-effectiveness analysis argues strongly for persistence in questioning and attempting to advance its theoretical foundations.

References

1. Udvarhelyi IS, Colditz GA, Rai A, Epstein AM. Cost-effectiveness and cost-benefit analyses in the medical literature: are the methods being used correctly? *Ann Intern Med.* 1992; 116:238-44.
2. Gold MR, Siegel JR, Russell LB, Weinstein MC (eds). *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press, 1996.
3. Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC, for the Panel on Cost-Effectiveness in Health and Medicine. The role of cost-effectiveness analysis in health and medicine. *JAMA.* 1996;276:1172-7.
4. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB, for the Panel on Cost-Effectiveness in Health and Medicine. Recommendations of the Panel on Cost-Effectiveness in Health and Medicine. *JAMA.* 1996;276:1253-8.
5. Siegel JE, Weinstein MC, Russell LB, Gold MR, for the Panel on Cost-Effectiveness in Health and Medicine. Recommendations for reporting cost-effectiveness analyses. *JAMA.* 1996;276:1339-41.
6. Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang Q. Multiattribute utility function for a comprehensive health status classification system: health utilities index mark 2. *Med Care.* 1996;34:702-22.
7. Brouwer WBF, Koopmanschap MA, Rutten FFH. Productivity costs measurement through quality of life? A response to the recommendations of the Washington Panel. *Health Econ.* 1997;6:253-9.
8. Weinstein MC, Siegel J, Garber A, et al. Productivity costs, time costs and health-related quality of life: a response to the Erasmus group. *Health Econ.* 1997;6:505-10.
9. Brouwer WBF, Koopmanschap MA, Rutten FFH. Productivity costs in cost-effectiveness analysis: numerator or denominator: a further discussion. *Health Econ.* 1997;6:511-14.
10. U.S. Bureau of the Census. *Statistical Abstract of the United States 1995*. Washington, DC: 1995:113.
11. Johannesson M. The concept of cost in the economic evaluation of health care: a theoretical inquiry. *Int J Technol Assess Health Care.* 1994;10:675-82.
12. Meltzer D. Accounting for future costs in medical cost-effectiveness analysis. *J Health Econ.* 1997;16(1): 33-64.
13. Weinstein MC, Manning W. Theoretical issues in cost-effectiveness analysis. *J Health Econ.* 1997;16(1):121-8.
14. Johannesson M, Meltzer D, O'Connor RM. Incorporating future costs in medical cost-effectiveness analysis: implications for the cost-effectiveness of the treatment of hypertension. *Med Decis Making.* 1997;17:382-9.
15. Garber A, Phelps C. Economic foundations of cost-effectiveness analysis. *J Health Econ.* 1997;16(1):1-32.

Appendix 7:

**On the Role of Theory in Cost-Effectiveness Analysis: A Response to Garber,
Russell, and Weinstein**

David Meltzer, Magnus Johannesson

Medical Decision Making

19:383-384,1999

Response

• On the Role of Theory in Cost-Effectiveness Analysis— A Response to Garber, Russell, and Weinstein

DAVID MELTZER, MD, PhD, MAGNUS JOHANNESSON, PhD

The comments of Garber, Russell, and Weinstein in response to our paper add to the important contributions to cost-effectiveness analysis that have grown out of the work of the Panel on Cost-Effectiveness Analysis in Health and Medicine.¹⁻³ Although they differ from each other and from our work in some areas, there appear to us to be many areas of agreement and important opportunities to further advance the field that come from the exchange among us.

A first area of agreement is that, to the extent that welfare economics is to serve as a basis for cost-effectiveness analysis, all future medical and non-medical costs resulting from increased survival should be included and measures of quality of life and costs should be designed to ensure that all benefits and costs are neither omitted nor counted more than once. Indeed, the Panel's report emphasizes the value of welfare economics as a "logical" foundation for cost-effectiveness analysis.⁴ We agree with this viewpoint, because welfare economics is based fundamentally on the principle of using limited societal resources to maximize social welfare and because it encourages the explicit consideration of both objectives and constraints. A second area of agreement is that a variety of practical and ethical concerns were, and should have been, factored into the Panel's recommendations about these issues, so that consistency with welfare economics should not be the sole consideration by which the Panel's recommendations are assessed. The heart of the issues raised in our exchange is about balancing these theoretical, practical, and ethical concerns.

With regard to the inclusion of productivity and other costs as part of quality-of-life measures, we

appear to be in broad agreement that there is ambiguity about costs in all existing quality-of-life measures (except the HUI). Moreover, we appear to agree that, despite the temptation on practical grounds to use existing measures without alteration, this ambiguity requires clarification both in posing quality-of-life questions and in measuring costs to ensure that all quantitatively important costs are neither omitted nor counted more than once. While we agree with the need to better quantify the magnitude of productivity losses not captured by present practices, we remain convinced that the substantial payroll and income taxes, as well as public or private disability insurance in most countries, make it extremely likely that such omissions are often a large part of total productivity losses. For this reason, we are pleased that at least two of our respondents agree with us that these concerns will likely be adequately addressed only by using measures that explicitly instruct respondents not to consider the personal economic consequences of their illness and to include the full productivity costs of morbidity in the numerator of the CE ratio. Testing and validating quality-of-life measures that meet these criteria is an important area for future work.

Only Garber and Weinstein respond specifically to the future-costs issue. Garber emphasizes the importance of measuring all future health and non-health costs unless they are small; we again note that the effects of including these costs will be small only under rather strict theoretical conditions and that actual calculations of the effects of including future costs show that such costs often cause substantial changes in cost-effectiveness ratios.^{5,6} We should also mention in response to Garber that the theoretical result about the need to include future costs that we illustrate with our example does not require the discontinuities in our simplified example; the same result is present in the continuous model analyzed in Meltzer's original paper on future costs.⁵

Weinstein objects to the inclusion of future non-medical costs of life extension on the grounds that this will reflect the current distribution of income and wealth in society and thus "be laden with some of the most objectionable features of CBA which im-

Received from the Section of General Internal Medicine, Department of Economics, and Harris Graduate School of Public Policy Studies, University of Chicago, Chicago, Illinois (DM); and the Centre for Health Economics, Stockholm School of Economics, Stockholm, Sweden (MJ). Accepted for publication July 21, 1999. Supported in part by a grant from the National Institute of Aging and The Robert Wood Johnson Foundation (Meltzer).

Address correspondence and reprint requests to Dr. Meltzer: Section of General Internal Medicine, University of Chicago, 5841 South Maryland MC 2007, Chicago, IL 60637; phone: (773) 702-0836; fax: (773) 834-2238; e-mail: (dmeltzer@medicine.bsd.uchicago.edu).

pede its more widespread acceptance." We note in response that, as Weinstein recognizes, including such costs is conceptually no different than using wages to value caregiver time, which the Panel endorses, and moreover that there is no need to use an individual's personal wage rate in such calculations, since one can use a population average if one wants differences in the distribution of wealth not to affect such calculations. In principle, distributional concerns could also be incorporated into cost-effectiveness analysis by weighting costs and QALYs differently in different population groups. It is, however, hard to see the logic of any weighting scheme that would inherently value a dollar spent due to extension of life differently than a dollar spent any other way. Moreover, we note that omission of future costs results in spending too much on interventions that extend life relative to those that improve quality of life in the elderly. In conclusion, we find the arguments for excluding future costs on

ethical or practical grounds unconvincing and recommend the inclusion of the full future costs in cost-effectiveness analysis.

References

1. Garber A. Response to David Meltzer and Magnus Johannesson. *Med Decis Making*. 1999;19:378-9.
2. Russell LB. Improving the Panel's recommendations—comment on Meltzer and Johannesson. *Med Decis Making*. 1999;19:379-80.
3. Theoretically correct cost-effectiveness analysis—a response to Meltzer and Johannesson. *Med Decis Making*. 1999;19:381-82.
4. Gold MR, Siegel JR, Russell LB, Weinstein MC (eds). *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press, 1996.
5. Meltzer D. Accounting for future costs in medical cost-effectiveness analysis. *J Health Econ*. 1997;16:33-64.
6. Johannesson M, Meltzer D, O'Connor RM. Incorporating future costs in medical cost-effectiveness analysis: implications for the cost-effectiveness of the treatment of hypertension. *Med Decis Making*. 1997;17:382-9.

Appendix 8:

Effect of Future Costs on the Cost-Effectiveness of Life Extension and Quality of Life Improvement among the Elderly

D Meltzer, G Hazen, M Johansson, I Abdalla, R Chang, A Elstein, A Schwartz

Medical Decision Making

18:475, October 1998

EFFECT OF FUTURE COSTS ON THE COST-EFFECTIVENESS OF LIFE EXTENSION AND QUALITY OF LIFE IMPROVEMENT AMONG THE ELDERLY. D Meltzer¹, G Hazen², M Johansson³, I Abdalla¹, R Chang³, A Elstein³, A Schwartz³. ¹U Chicago, ²Northwestern U, ³U Illinois, ⁴Stockholm Sch. Econ.

Purpose: The inclusion of future medical costs for "unrelated" illnesses and future non-medical costs in cost-effectiveness (C/E) analyses is controversial. Recent theoretical work suggests that these costs should be included in C/E analyses and that failure to include them will bias analyses to favor interventions that extend life expectancy (LE) over interventions that improve quality of life (QOL). However, some analysts have been uncomfortable including future costs because they think that this would prevent all interventions in the elderly from being cost-effective. To assess this, we contrast the effects of including future costs on the C/E of three common medical interventions among the elderly that differ in their effects on QOL versus LE: hip replacement for osteoarthritis, treatment of hypertension, and treatment for prostate cancer.

Methods: We modified published decision and C/E analyses to include future medical and non-medical costs and calculated the cost per QALY. Due to uncertainty about the C/E ratio based on current and related costs (CRC) of treating prostate cancer, we report those estimates contingent on CRC.

Results:

INTERVENTION	Δ LE	Δ QALY	Δ LE/ Δ QALY	Δ Cost/ Δ QALY w/o future costs	Δ Cost/ Δ QALY w/ future costs
Hip Replacement, women age 60	-0.03	6.9	-0.005	Cost-saving	Cost-saving
Hip Replacement, men age 85	-0.02	2.0	-0.01	\$9,177	\$9,042
Treatment Mild HTN, men age>70	0.06	0.05	1.25	\$5,000	\$32,000
Radiation Therapy MDPC, age 65	0.8	0.4	2	CRC	CRC+\$32,000
Prostatectomy MDPC, age 65	0.7	0.2	3.5	CRC	CRC+\$56,000

HTN=Hypertension. MDPC=Moderately differentiated prostate cancer. CRC=Current + related costs.

Conclusions: Including future costs does not prevent interventions for the elderly from being cost-effective. However, including future costs does improve the C/E of interventions that improve QOL relative to those that increase LE. Some interventions with small gains in QALY compared to LE may not be C/E even if CRC are zero. Resource allocations based on analyses that exclude future costs may spend too much on interventions that extend life compared to interventions that improve QOL. The effects of including future costs should be reported in all C/E analyses.

Appendix 9:

Do Quality-Adjusted Life Years Reflect Patient Preferences? Validation Using Revealed Preference for Intensive Treatment of Insulin-Dependent Diabetes

Mellitus

D Meltzer, T Polonsky

Medical Decision Making

18:459, October 1998

DO QUALITY-ADJUSTED LIFE YEARS REFLECT PATIENT PREFERENCES? VALIDATION USING REVEALED PREFERENCE FOR INTENSIVE TREATMENT OF INSULIN-DEPENDENT DIABETES MELLITUS. DO Meltzer, TS Polonsky. University of Chicago, Chicago, IL.

Purpose: Quality-adjusted life years (QALYs) are often used to evaluate therapies in medical cost-effectiveness analysis, but their validity as a measure of patient preferences is unproven. The absence of an accepted measure of patient preferences has hindered efforts to validate QALYs. We used the clinically accepted standard for identifying the preferences of patients – their informed choices – to assess the "revealed" preferences of 130 patients with IDDM for intensive (INT) vs. conventional (CONV) therapy. To assess the validity of QALYs as a measure of patient preferences, we then used standard techniques to calculate the change in QALYs with INT and examined whether this predicts patient choice of therapy.

Methods: Time trade-off questions were used to measure patient beliefs about quality of life (QOL) with each therapy and the major complications of IDDM. Frequency assessment questions were used to assess patient beliefs about the efficacy of INT. Questions about preference for the present vs. future were used to assess discount rates. A decision model based on the Diabetes Control and Complications Trial (DCCT) was used to calculate the change in QALYs with INT.

Results: Patients on CONV reported lower QOL with INT vs. CONV ($p < 0.01$), while patients on INT reported no difference in QOL. However, patients on INT and CONV did not differ in the QOL they associated with complications of IDDM. Patients on INT reported higher estimates of efficacy of INT in preventing complications than did patients on CONV ($p < 0.01$), but overestimated the benefits of INT compared to the results of the DCCT. Patients on INT also exhibited lower discount rates than did patients on CONV. The change in QALYs predicted reported treatment, with an area under the ROC curve (AUC) of 0.84. Change in QALYs also predicted the behaviors that define INT, but the AUC was only 0.73, perhaps because the correlation of QALYs and reported therapy is enhanced by attempts to minimize cognitive dissonance.

Conclusions: QALYs correlate with patient preferences as revealed by their choices. However, the possibility that answers reflect patient efforts to minimize cognitive dissonance concerning their choice of therapy suggests that this should be considered an upper bound on the extent to which QALYs reflect patient preferences. Analyses whose results depend on measurement of QALYs should be interpreted with caution. Revealed preference techniques should be used to validate and improve QOL assessment.

Appendix 10:

Structure and Early Results of a Prostate Cancer Decision Model

D Meltzer, A Basu, B Egleston

Medical Decision Making

In Press

Structure and Early Results of a Prostate Cancer Decision Model

David O. Meltzer, Anirban Basu, Brian Egleston, University of Chicago, Chicago, IL.

Purpose: To develop an innovative decision model to assess the effectiveness and cost-effectiveness of alternate treatment strategies on men with clinically localized prostate cancer (PC).

Methods: We reviewed all existing decision models of PC screening and/or treatment. We then developed a new model based on pathologic stage and grade of PC that addresses many important aspects of the natural history of prostate cancer progression neglected by previous analysis. Key features include: modeling of tumor heterogeneity and progression rates by grade, misclassification of tumor stage due to discrepancy between clinical and pathological stage, and ability of screening and treatment to affect prevalence of PC in the undetected population, and accounting for the effect of benign prostatic hypertrophy and its management on the detection of prostate cancer. The core of the model consist of a Markov process defined by a state vector containing 1440 elements spanning durable characteristics of PC such as detectability. The model is then run from age 50 to 99 years to assess the effects of screening and treatment policies. Model parameters are derived by extensive literature review and validated using aggregate incidence and mortality data as well as individual level data using maximum likelihood techniques. In this abstract we report early analyses of the effects of alternative treatments on outcomes by tumor grade (well, moderate and poorly differentiated) for men 50 to 80 years of age.

Results: Treatment offers only modest increase (maximum 0.5 years) in non-discounted life expectancy for well-differentiated tumors at all ages. Effects are more favorable (> 1 year) for moderately and poorly differentiated tumors, but only for men below age 70. Treatment was cost-effective for men in this category (CER \$40K - \$76K per non-discounted life year). When quality of life weights were added, quality-adjusted life expectancy (QALE) was reduced. QALE increased only in extreme sensitivity analyses that ignored age-specific decrements in QOL, complications of treatment, or discounting. Even in these extreme cases, treatment was at best marginally cost-effective by the \$100,000 per QALY standard.

Conclusions: The preliminary results of our model suggest that radical prostatectomy and radiation therapy improve non-discounted life expectancy for younger patients with more advanced tumors at a cost of less than \$100,000 per life year saved. Similar benefits are not evident in QALE, but the results may be sensitive to assumptions about discount rates or QOL weights.